Probab. Th. Rel. Fields 85, 147–180 (1990)

Probability Theory and Related Fields

Nonparametric High Resolution Spectral Estimation*

R. Dahlhaus

Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, D-6900 Heidelberg, Federal Republic of Germany

Received November 2, 1988; in revised form December 27, 1989

Summary. The uniform rate of convergence of the integrated relative mean square error over a (with the sample size T) increasing class $\mathscr{X}_T$ of stationary processes is studied for several estimates of the spectral density. The class $\mathscr{X}_T$ is chosen in a way such that estimates with a good uniform rate of convergence over $\mathscr{X}_T$ may be termed 'high resolution spectral estimates'. By using this criterion several effects are explained theoretically, for example the leakage effect. The advantages uf using data tapers are proved and the use of local and global bandwiths are studied. Furthermore, the behavior of segment estimates are studied. Simulations are presented for the illustration of some effects.

Contents

1. Introduction

This paper is concerned with the nonparametric estimation of the spectral density $f(\lambda)$ of a stationary process X_t, $t \in \mathbf{Z}$ from the sample $X_0, \ldots, X_{T-1}$. The estimation

* This work has been supported by the Deutsche Forschungsgemeinschaft

usually is done by calculating the periodogram which, in the case of a zero mean, has the form

$$I_T(\lambda)=\{2\pi H_{2,T}\}^{-1}\left|\sum_{t=0}^{T-1} h_{t,T}X_t\exp(-i\lambda t)\right|^2$$

where

$$H_{k,T}=\sum_{t=0}^{T-1} h_{t,T}^k, \qquad k\in\mathbb{N}_0 .$$

$h_{t,T}$ is a data taper, e.g. the cosine bell $h_{t,T}=\frac{1}{2}\,[1-\cos\{2\pi(t+0.5)/T\}]$.

Since the periodogram is only asymptotically unbiased but not consistent it has to be smoothed. One usually considers the estimate

$$f_T^{(1)}(\lambda)=\int_\Pi I_T(\lambda+\alpha)\,W_N(\alpha)\,d\alpha \quad \Pi:=(-\pi,\pi] \tag{1.1}$$

with a suitable kernel $W_N(\alpha)$ e.g. $W_N(\alpha)=NW(N\alpha)$ with $\frac{N}{T}\to 0$.

Although the usual asymptotic theory (asymptotic normality, integrated mean square error, etc.) leads to satisfactory results for the estimate, $f_T^{(1)}(\lambda)$ behaves in certain situations rather badly. Several negative effects may arise that could not be explained successfully by the mathematical theory so far.

Problems arise for example if strong peaks are present in the spectrum. If no data taper is used ($h_{t,T}\equiv 1$) the estimate is not able to resolve lower peaks of the spectrum. This effect has been called leakage effect. It can be cured by application of a data taper (cf. Bloomfield, 1976, Sect. 5.2). To illustrate the effect we have plotted in Fig. 1 the true spectrum of an (AR(14)-process (dark line) the kernel estimate (1.1) (more precisely (3.9)) with the nontapered periodogram and a global bandwith (dotted line) and the same estimate with a tapered periodogram (dashed line) (for details of the simulation see Sect. 4.1). We clearly see the strong bias of the nontapered estimate. Although this effect has been known for a long time, it has never been described theoretically in a stringent way. The ordinary asymptotic theory only shows disadvantages of data tapers: the variance and the mean square error of the estimate increase with the use of a taper (cf. Brillinger, 1981, Theor. 5.6.4).

As a consequence of the bad behaviour of the above nontapered estimate (and other nonparametric estimates as well) applied workers (especially engineers) very often prefer a parametric (usually AR-) approach together with estimation procedures that have high resolution properties, e.g. the maximum entropy method (Burg-algorithm). Such procedures are termed 'high resolution spectral estimates' (cp. the articles in Childers, 1978).

In this paper we will make an attempt to define by a mathematical model what is meant by 'high resolution spectral estimates', and to explain theoretically the leakage effect and other effects that may arise in nonparametric spectral estimation. Since most of the effects are small sample effects which disappear asymptotically we create a special asymptotic model by allowing e.g. the peaks to increase with the sample size.

Since the variance of a spectral estimate f_T is usually proportional to f it is natural to consider as a measure of goodness of an estimate the integrated relative mean square error.

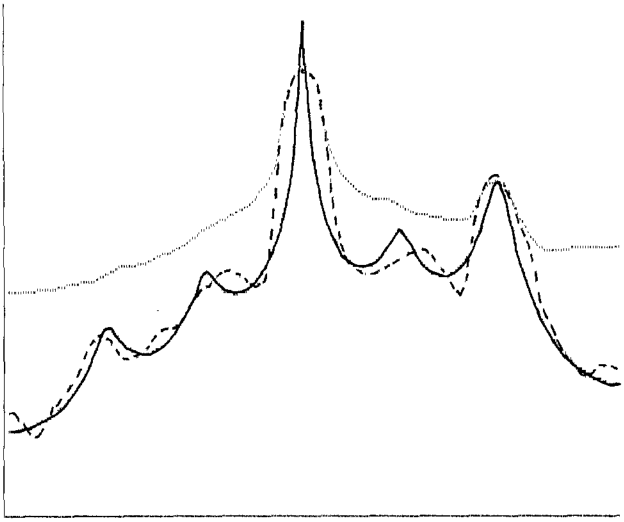

Fig. 1. Window estimate with and without data taper

Let

$$\mathrm{IBIAS}(f_T) = \int_{-\pi}^{\pi} \left(\frac{Ef_T(\lambda)}{f(\lambda)} - 1\right)^2 d\lambda ,$$

$$\mathrm{IVAR}(f_T) = \int_{-\pi}^{\pi} \mathrm{var}\left(\frac{f_T(\lambda)}{f(\lambda)}\right) d\lambda$$

$$\mathrm{IMSE}(f_T) = \int_{-\pi}^{\pi} E\left(\frac{f_T(\lambda)}{f(\lambda)} - 1\right)^2 d\lambda ,$$

and SBIAS(f_T), SVAR(f_T), SMSE(f_T) be the corresponding statistics with the integral replaced by the sum $\frac{2\pi}{T} \sum_{\substack{\lambda=\lambda_S \\ s=1,\ldots,T}}$ where $\lambda_S = \frac{2\pi s}{T}$.

We will study the convergence rate of

$$\sup_{\mathscr{X}_T} \mathrm{IMSE}(f_T) \tag{1.2}$$

for several estimates where $\mathscr{X}_T$ is a (with T) increasing class of stochastic processes. By using an increasing class we require that the estimates behave uniformly good over an increasing number of stochastic processes when the sample size increases. By using this model we avoid that certain small sample effects such as the leakage effect disappear asymptotically. The class $\mathscr{X}_T$ is defined in Sect. 2. It contains processes with spectral densities that have with T increasing peaks and troughs, for example autoregressive moving average processes with characteristic roots up to T^{-1} close to the unit circle. By this choice of $\mathscr{X}_T$ we are able to cover asymptotically problem cases in statistical inference. In particular, we are able to discuss the resolution properties of the estimates. Estimates f_T with $\sup_{\mathscr{X}_T} \mathrm{IMSE}(f_T) = O(T^{-4/5})$ will be termed 'high resolution spectral estimates'. $T^{-4/5}$ is the usual rate of convergence of the IMSE for window estimates with a positive kernel.

In Sect. 3 we prove that tapered window estimates with a certain local bandwith have this high resolution property.

In Sect. 4 we present several window estimates that have a lower uniform rate of convergence, among them estimates with a global bandwith. For nontapered estimates we prove that $\sup_{\mathscr{X}_T}$ IMSE (f_T) does not even converge to zero. This explains theoretically the leakage effect. The same holds for nontapered estimates when the spectrum contains troughs. We call this effect 'trough effect'. This is the first time that the trough effect is described. Furthermore, we prove that tapering may not only reduce the bias but also the variance of window estimates, which is contrary to widespread conjectures.

Some effects are demonstrated by simulations.

In Sect. 5 we consider segment estimates, i.e. estimates obtained by averaging periodograms over overlapping data segments. We prove that these estimates also have a lower uniform rate of convergence.

The proofs are very technical. In order to make the paper more readable we have put nearly all proofs into the appendix.

In Dahlhaus (1988) we have derived similar results for parametric estimates.

A key role in our calculations is played by the following function. Let L_T: $\mathbb{R}\to\mathbb{R}$, $T\in\mathbb{R}^+$, be the periodic extension (with period 2π) of

$$L_T^*(\alpha)=\begin{cases} T, & |\alpha|\leq 1/T \\ 1/|\alpha|, & 1/T<|\alpha|\leq\pi \end{cases}. \tag{1.3}$$

The function $L_T(\alpha)$ is used to describe the properties of data tapers, to define the class $\mathscr{X}_T$ and as a tool for handling the cumulants of time-series statistics. The properties of $L_T(\alpha)$ are summarized in Lemma A.1.

We further use cumulants and cumulant spectra of stationary processes. For the definitions and the basic properties we refer to Brillinger (1981, Sect. 2.3, especially Theor. 2.3.2).

2. The High Resolution Property

We now define the class of processes $\mathscr{X}_T$ over which the estimates are supposed to be uniformly good.

(2.1) Definition. Let $s_1, s_2\in\mathbb{N}_0$, $T\in\mathbb{N}$ and $\delta_0>0$, $C_0>1$. By $\mathscr{X}(T,s_1,s_2,\delta_0,C_0)$ we denote the set of all fourth order stationary processes X_t, $t\in\mathbb{Z}$ that can be represented in the form

$$X_t=\sum_{s=-\infty}^{\infty} a_s Y_{t-s}$$

where

(i) Y_t is a fourth order stationary process Y_t with mean 0, three times times differentiable spectrum $f_Y=f_{2,Y}$ with $C_0^{-1}\leq f_{2,Y}\leq C_0$, $f_{2,Y}^{(k)}\leq C_0$ $(k=1,2,3)$, and continuous fourth order spectrum $f_{4,Y}$ with $f_{4,Y}\leq C_0^2$.

(ii) The transfer function $A(\lambda)=\sum_{s=-\infty}^{\infty} a_s\exp(-i\lambda s)$ is of the form

$$A(\lambda)=\frac{\prod_{j=1}^{r_1} A_{1j}(\lambda-\lambda_{1j})^{s_{1j}}}{\prod_{j=1}^{r_2} A_{2j}(\lambda-\lambda_{2j})^{s_{2j}}} \tag{2.1}$$

with $s_{ij} \leqq s_i (i=1,2)$ and $|\lambda_{i_1 j_1} - \lambda_{i_2 j_2}| \geqq 2\delta_0 \,(\mathrm{mod}\, 2\pi)$ for $(i_1, j_1) \neq (i_2, j_2)$ $\left(\text{i.e. } r_i \leqq \frac{\pi}{\delta_0}\right)$.

The $g_{ij} = |A_{ij}|^2$ are three times differentiable with

$$C_0^{-1} L_{T_{ij}}(\lambda)^2 \leqq g_{ij}(\lambda)^{-1} \leqq C_0 L_{T_{ij}}(\lambda)^2 \ (\lambda \in \Pi) , \tag{2.2}$$

where $T_{ij} \leqq T$, and

$$|g'_{ij}(\lambda)| \leqq C_0 |\lambda| \ (\lambda \in \Pi) , \tag{2.3}$$

$$|g''_{ij}(\lambda)| \leqq C_0 , \quad |g^{(3)}_{ij}(\lambda)| \leqq C_0 \ (\lambda \in \Pi) , \tag{2.4}$$

$$|g''_{ij}(\lambda)| \geqq C_0^{-1} \ (|\lambda| \leqq \delta_0) . \tag{2.5}$$

Note, that $\mathscr{X}(T, s_1, s_2, \delta_0, C_0)$ is monotone in all five variables. We sometimes drop the subscripts and denote $\mathscr{X}(T, s_1, s_2, \delta_0, C_0)$ by $\mathscr{X}_T$.

An estimate f_T with $\sup_{\mathscr{X}_T} \mathrm{IMSE}(f_T) = O(T^{-4/5})$ will be termed 'high resolution spectral estimate'.

(2.2) *Remarks.* (i) The conditions (2.2)–(2.5) could also be formulated in terms of the A_{ij}. Since the second order spectrum of the process $X_t \in \mathscr{X}_T$ is of the form

$$f(\lambda) = f_{2,X}(\lambda) = \frac{\prod_{j=1}^{r_1} g_{1j}(\lambda - \lambda_{1j})^{s_{1j}}}{\prod_{j=1}^{r_2} g_{2j}(\lambda - \lambda_{2j})^{s_{2j}}} f_{2,Y}(\lambda)$$

we chose the formulation in terms of the g_{ij}. An example for a g_{ij} that fulfills (2.2)–(2.5) is $g_{ij}(\lambda) = g(\lambda) = \frac{1}{T^2} + \lambda^2$. Thus, f has strong peaks of magnitude T_{2j}^2 and of multiplicity s_{2j} at frequencies $\lambda_{2j}\,(j=1,\ldots,r_2)$ and troughs of magnitude T_{1j}^{-2} and of multiplicity s_{1j} at frequencies $\lambda_{1j}\,(j=1,\ldots,r_1)$. Below we prove that the class $\mathscr{X}_T$ contains all ARMA-processes with roots up to $1/T$ close to the unit circle.

(ii) At first sight the conditions on the existence of the third derivatives seem to be inadequate for the treatise of strong peaks. However, the above conditions allow e.g. a value $\sim s_i T^2$ for $|(\log f)''|$ at a peak (trough) λ_{ij}.

(2.3) Theorem. *Let* $s_1, s_2, T \in \mathbb{N}_0$, $0 < \delta_0 < \frac{\pi}{3}$, $C_0 \geqq 20$, Y_t *be a stationary process that fulfills the conditions of Definition* 2.1(*i*) *and* X_t *be defined by*

$$\sum_{j=0}^{p} a_j X_{t-j} = \sum_{j=0}^{q} b_j Y_{t-j}$$

where

$$\sum_{j=0}^{p} a_j e^{i\lambda j} = \prod_{j=1}^{r_2} (1 - q_{2j} e^{-i\lambda_{2j}} e^{i\lambda})^{s_{2j}}$$

and

$$\sum_{j=0}^{q} b_j e^{i\lambda j} = \prod_{j=1}^{r_1} (1 - q_{1j} e^{-i\lambda_{1j}} e^{i\lambda})^{s_{1j}}$$

with

$C_0^{-1} < q_{ij} \leqq 1 - 1/T$, $|\lambda_{i_1 j_1} - \lambda_{i_2 j_2}| \geqq 2\delta_0 \,(\mathrm{mod}\, 2\pi)$ *for* $(i_1, j_1) \neq (i_2, j_2)$, *and* $s_{ij} \leqq s_i$.

Then

$$X_t \in \mathscr{X}(T, s_1, s_2, \delta_0, C_0) .$$

Proof. The transfer function of X_t is of the form (2.1) with $A_{ij}(\lambda) = 1 - q_{ij}e^{i\lambda}$. Direct calculation gives

$$\frac{4}{\pi^2}\,[(1-z)^2 + z\lambda^2] \leqq |1 - ze^{i\lambda}|^2 \leqq (1-z)^2 + z\lambda^2$$

and we therefore obtain with $T_{ij} = (1 - q_{ij})^{-1}$

$$L_{T_{ij}}(\lambda)^2 \leqq g_{ij}(\lambda)^{-1} \leqq 2\pi^2 L_{T_{ij}}(\lambda)^2 . \tag{2.6}$$

Since

$$g_{ij}(\lambda) = 1 + q_{ij}^2 - 2q_{ij}\cos\lambda$$

also (2.3)–(2.5) are fulfilled.

We now derive an expansion for spectral densities of processes $X_T \in \mathscr{X}_T$. This is the main property of the class $\mathscr{X}_T$. Let $r_1, r_2, m_1, m_2 \in \mathbb{N}_0$ with $m_1 + m_2 \geqq 1$ and $r_3 = r_2$, $T \in \mathbb{N}$, $\underline{\lambda}_1 \in \Pi^{r_1}$, $\underline{\lambda}_2 \in \Pi^{r_2}$, $\underline{T}_1 \in \mathbb{R}_+^{r_1}$ and $\underline{T}_2 \in \mathbb{R}_+^{r_2}$. We define for $\ell \in \mathbb{N}_0$

$$R(\lambda, \alpha, \ell, m_1, m_2) = R(\lambda, \alpha, \ell, m_1, m_2, \underline{\lambda}_1, \underline{\lambda}_2, \underline{T}_1, \underline{T}_2, r_1, r_2)$$

$$= \sum_{\substack{j_i=1\\ i=1,2,3}}^{r_i} \Bigg[\sum_{\substack{k_1=0,\ldots,m_1\\ k_2=0,\ldots,m_2\\ k_3=0,\ldots,\max(0,\min(\ell-1,m_2))\\ \ell \leqq k_1+k_2+k_3}} |\alpha|^{k_1+k_2+k_3} L_{T_{1j_1}}(\lambda - \lambda_{1j_1})^{k_1} L_{T_{2j_2}}(\lambda + \alpha - \lambda_{2j_2})^{k_2}$$

$$\cdot L_{T_{2j_3}}(\lambda - \lambda_{2j_3})^{k_3} + \{m_1 = 2, m_2 = 0, \ell = 3\}\, |\alpha|^3 L_{T_{1j_1}}(\lambda - \lambda_{1j_1})^2 \Bigg] .$$

(2.4) Theorem. *Let $X_t \in \mathscr{X}(T, s_1, s_2, C_0, \delta_0)$ with spectral density f, transfer function A, and fourth order spectrum f_4. Then we have for $\ell \in \{0, 1, 2, 3\}$*

(a) $$\frac{f(\lambda+\alpha)}{f(\lambda)} = \sum_{k=0}^{\ell-1} \frac{f^{(k)}(\lambda)}{f(\lambda)} \frac{\alpha^k}{k!} + O[R(\lambda, \alpha, \ell, 2s_1, 2s_2)],$$

(b) $$\left|\frac{A(\lambda+\alpha)}{A(\lambda)}\right| - 1 = O[R(\lambda, \alpha, 1, s_1, s_2)]$$

and

(c) $$\frac{f_4(\gamma_1, \gamma_2, \gamma_3)}{f(\beta_1) f(\beta_2)} = O[R(\beta_1, \gamma_1 - \beta_1, 0, s_1, s_2) R(-\beta_1, \gamma_2 + \beta_1, 0, s_1, s_2) \cdot R(\beta_2, \gamma_3 - \beta_2, 0, s_1, s_2) R(-\beta_2, -\gamma_1 - \gamma_2 - \gamma_3 + \beta_2, 0, s_1, s_2)].$$

The O-terms depend only on s_1, s_2, C_0 and δ_0.

We should note that the remainder terms in the above expansions are usually *not* small. The remainder term is for example large in (a) if f has a strong peak at $\lambda + \alpha$ and is smooth at λ.

3. High Resolution Window Estimates

In this section we prove that the window estimate (1.1) with a suitable data taper and a suitable local bandwith selection is a high resolution estimate.

3.1 Data Tapers

As we will show a good taper is for example of the form $h_{t,T}=h\left(\frac{t}{T}\right)$ where h is sufficiently smooth, expecially at 0 and 1 (which is not fulfilled in the nontapered case $h(x)=I_{[0,1)}(x)$). This smoothness is determined by the 'degree' of the taper defined below.

(3.1) Definition. Let $k\in\mathbb{N}_0$ and $\kappa\in[0,1/2)$. Suppose $h_{t,T}=h_T\left(\frac{t}{T}\right)$ is a sequence of data tapers with $h_T(x)=0$ for all $x\notin[0,1)$ that fulfills the following conditions.

(i) h_T is $(k-1)$-times continuously differentialbe (in the case $k=1$ we assume continuity and in the case $k=0$ we make no assumption).

(ii) There exists a finite set $P_T=\{p_{1T},\ldots,p_{rT})$ such that h_T is $(k+1)$-times differentiable in all $x\notin P_T$.

(iii) Let $s_{jT}:=\lim\limits_{y\downarrow p_{jT}} h_T^{(k)}(y)-\lim\limits_{y\uparrow p_{jT}} h_T^{(k)}(y)$. There exists a $c>0$ such that $\sum_{j=1}^{r} s_{jT}^2\geqq c$ for all $T\in\mathbb{N}$.

(iv) $H_{2,T}\sim T$ and $D_T^{(k)}:=\sup\limits_{x\notin P_T}|h_T^{(k)}(x)|+\sup\limits_{x\notin P_T}|h_T^{(k+1)}(x)|\leqq kT^{\kappa}$ with $\kappa\in[0,1/2)$.

Then we say that the taper (the sequence of tapers) is of degree (k,κ).

In the non-tapered case $h_{t,T}=\chi_{[0,1)}(t/T)$ the degree therefore is $(0,0)$.

(3.2) *Example (Polynomial Taper).* The function

$$h_\varrho(x)=\begin{cases}4^k(x/\varrho)^k(1-x/\varrho)^k, & x\in[0,\varrho/2)\\ 1, & x\in[\varrho/2,1/2]\\ h_\varrho(1-x), & x\in(1/2,1]\end{cases}$$

is $(k-1)$-times continuously differentiable and $(k+1)$-times differentiable in $x\notin P=\{0,\varrho/2,1-\varrho/2,1\}$. Thus, the taper $h_{t,T}=h_\varrho\left(\frac{t}{T}\right)$ where ϱ is fixed has degree $(k,0)$. Furthermore, we have $\sup\limits_{x\notin P}|h_\varrho^{(\ell)}(x)|\leqq K\varrho^{-\ell}(0\leqq\ell\leqq 2k)$ with K independent of ϱ. Thus, if e.g. $\varrho=\varrho_T=T^{-\kappa/(k+1)}$ the taper $h_{t,T}=h_{\varrho_T}\left(\frac{t}{T}\right)$ has degree (k,κ) with $h_{\varrho_T}(x)\to\chi_{(0,1)}(x)$ and $\lim\limits_{T\to\infty}\frac{TH_{4,T}}{H_{2,T}^2}=1$ (cp. the discussion below Theorem 3.6).

Another example is the Tukey-Hamming taper (cp. Dahlhaus, 1988, Ex. 5.2).

We now state a fundamental inequality for data tapers. Let

$$H_k^{(T)}(\lambda)=\sum_{t=0}^{T-1} h_{t,T}^k\exp(-i\lambda t)\quad\text{and}\quad H_T(\alpha)=H_1^{(T)}(\alpha)\,.$$

(3.3) Lemma. *Let $k\in\mathbb{N}_0$, $\kappa\in[0,1/2)$ and $(h_{t,T})_{T\in\mathbb{N}}$ be a sequence of data tapers of degree (k,κ). Then there exists a constant $K\in\mathbb{R}$ such that for all $\alpha\in\mathbb{R}$ and $T\in\mathbb{N}$*

$$|H_T(\alpha)|/H_{2,T}^{1/2}\leqq KT^{-k-1/2+\kappa}L_T(\alpha)^{k+1}\,. \tag{3.1}$$

Proof. The lemma is proved in Dahlhaus (1988, Lemma 5.4).

In practical situations we do not want to drop the first observation X_0 completely which would happen by using the taper $h_{t,T}=h_T(t/T)$ with $h_T(0)=0$. One therefore chooses in practice the taper $h_{t,T}=h_T((t+1/2)/T)$ which also fulfills (3.1).

3.2 Local Bandwith Selection

Another ingredient of high resolution window estimation is the bandwith selection. The usual arguments for bandwith selection are as follows. Under suitable regularity conditions one obtains

$$\frac{E\hat{f}_T(\lambda)}{f(\lambda)}-1=N^{-2}\,\frac{f''(\lambda)}{f(\lambda)}\cdot\frac{1}{2}\int \alpha^2\, W(\alpha)\,d\alpha+o(N^{-2}) \tag{3.2}$$

and

$$\operatorname{var}\frac{\hat{f}_T(\lambda)}{f(\lambda)}=\frac{N}{T}\,2\pi\int W(\alpha)^2\,d\alpha+o\left(\frac{N}{T}\right). \tag{3.3}$$

Minimizing the relative mean square error with respect to N then leads to the optimal (local) bandwith

$$B_T:=N^{-1}$$

where

$$N=c_W\,T^{1/5}\left|\frac{f''(\lambda)}{f(\lambda)}\right|^{2/5}. \tag{3.4}$$

with a certain constant c_W depending on $W(\alpha)$. If $f''(\lambda)=0$ one has to make a higher order expansion for the bias which leads to a lower rate than $T^{1/5}$ for N.

In the class $\mathscr{X}_T$ introduced in Sect. 2 the bandwith selection is more difficult. Firstly, $\frac{f''(\lambda)}{f(\lambda)}$ may increase with T. From (2.2) to (2.4) we see that $\frac{f''(\lambda)}{f(\lambda)}$ may take at sharp peaks the value T^2. The above Definition (3.4) then would lead to $N\tilde{\ }T$ and a bandwith $B_T\tilde{\ }T^{-1}$ which is in accordance with our intuition. However, this causes considerable technical problems since the usual assumption $B_T T\to\infty$ is violated. Secondly, the expansion (3.2) for the bias is only good if $\frac{f''(\lambda)}{f(\lambda)}$ is approximately the same over the whole range of the bandwith. In the class $\mathscr{X}_T$ this is not true. Consider for example an AR(1) process with root $p_T=1-1/T$ and spectral density

$$f(\lambda)=\frac{1}{2\pi}\,|1-p_T e^{i\lambda}|^{-2}\,.$$

Elementary calculations give $\frac{f''(0)}{f(0)}=-2T(T-1)$ and $\frac{f''(\lambda)}{f(\lambda)}=0$ for $|\lambda|\approx 3^{-1/2}T^{-1}$. This means that the maximum and the point of inflection are less than T^{-1} apart (less than one Fourier frequency!) while the bandwith selection (3.4) would lead to $N\tilde{\ }T(B_T\tilde{\ }T^{-1})$ at $\lambda=0$ (which seems to be reasonable) and to $N=0$ ($B_T=\infty$; practically to a large bandwith) at $\lambda=3^{-1/2}T^{-1}$ which clearly is a bad choice.

Thus, we have to use a bandwith B_T which is not too large in the neighbourhood of strong peaks. One way to guarantee this is to incorporate the first derivative into the bandwith selection. A bandwith that fits our needs is

$$B_T = N^{-1} \quad \text{with}$$
$$N = N(T, f, \lambda) = cT^{1/5}\{|(\log f)''| + (\log f)'^2 + 1\}^{2/5} \tag{3.5}$$

with an arbitrary fixed constant $c > 0$. Since $\frac{f''(\lambda)}{f(\lambda)} = (\log f)'' + (\log f)'^2$, this bandwith is close to the bandwith (3.4) if $(\log f)'' > 0$.

For the derivation of the high resolution property of the window estimate we do not need the special form (3.5) but only the following two properties of this bandwith.

(3.4) Lemma. *There exist constants C_1, C_2 only depending on s_1, s_2, C_0 and δ_0 such that we have for the bandwith selection* (3.5)

$$N = N(T, f, \lambda) \leqq C_1 T^{1/5} \left(\sum_{i=1}^{2} \sum_{j=1}^{r_i} L_{T_{ij}}(\lambda - \lambda_{ij})^2 \right)^{2/5} \tag{3.6}$$

and

$$N = N(T, f, \lambda) \geqq C_2 T^{1/5} \max_{i,j} L_{T_{ij}}(\lambda - \lambda_{ij})^{4/5} \tag{3.7}$$

uniformly for all spectral densities f_X of processes $X_t \in \mathscr{X}(T, s_1, s_2, \delta_0, C_0)$ (with T_{ij} and λ_{ij} as in Def. 2.1).

(3.5) *Remark.* Since the bandwith (3.5) depends on the unknown spectral density this bandwith selection is not very helpful in practice. In our simulation (cp. Sect. 4.1) we have used this bandwith where f was replaced by a preliminary estimate. In order to get a high resolution estimate we only need the properties (3.6) and (3.7). Therefore, any other estimate with these properties would do. It would be very interesting to know whether practical suggestions such as using a bootstrap technique for bandwith selection (cp. Franke and Härdle, 1988) will lead to a high resolution estimate in the sense of this paper.

3.3 Convergence of High Resolution Window Estimates

We now discuss the properties of the window estimate $f_T^{(1)}$ (cp. (1.1)), where

$$W_N(\alpha) = NW(N\alpha) \text{ with a function } W: \mathbb{R} \to [0, c_1]\ (c_1 > 0) \text{ of} \tag{3.8}$$

$$\text{bounded variation with } W(\beta) = 0 \text{ for } |\beta| > c_2 \text{ and } \int_{-c_2}^{c_2} W(\beta)\, d\beta = 1\,.$$

Since the periodogram can be calculated rapidly at the frequencies $\lambda = \lambda_s$ with the Fast Fourier Algorithm one prefers in practice the estimate

$$f_T^{(2)}(\lambda) = \frac{2\pi}{T} \sum_{s=0}^{T-1} I_T(\lambda + \alpha_s)\, W_N(\alpha_s)\,, \qquad \alpha_s = \frac{2\pi s}{T} \tag{3.9}$$

($\lambda + \lambda_s = 0$ is often excluded in the sum – this would not afflict our results). However, due to the formula

$$f_T^{(1)}(\lambda)=\frac{1}{2\pi}\sum_{u=-(T-1)}^{T-1} c_T(u)\,w\left(\frac{u}{N}\right)\exp(-i\lambda u)$$

with the empirical tapered covariances

$$c_T(u)=H_{2,T}^{-1}\sum_{0\leq t,t+u\leq T-1} h_{t,T}X_t h_{t+u,T}X_{t+u}$$

and

$$w(u)=\int_{-\pi}^{\pi} W(\alpha)\exp(i\alpha u)\,d\alpha$$

the estimate $f_T^{(1)}(\lambda)$ can be calculated exactly.

We now prove that $f_T^{(1)}$ and $f_T^{(2)}$ are high resolution estimates, i.e. that (1.2) has rate of convergence $T^{-4/5}$.

(3.6) Theorem. *Let* $s_1, s_2 \in \mathbb{N}_0$, $s_1+s_2\geq 1$, $\delta_0>0$, $C_0>1$, *and* $(h_{t,T})_{T\in\mathbb{N}}$ *be a sequence of data tapers of degree* (k,κ) *with* $k\geq \max\{s_1, s_2, s_1+s_2-1\}$ *and* $\kappa<1/40$. *Furthermore, let* $N=N(T,f,\lambda)$ *fulfill* (3.6) *and* (3.7) (*take* e.g. (3.5)). *Then we have for* $i=1,2$

(a) $$\sup_{\mathscr{X}_T} \operatorname{IBIAS}(f_T^{(i)})=\frac{1}{4}\left(\int_{-c_2}^{c_2}\alpha^2 W(\alpha)\,d\alpha\right)^2 \sup_{\mathscr{X}_T}\int_{-\pi}^{\pi}\frac{1}{N^4}\left|\frac{f''(\lambda)}{f(\lambda)}\right| d\lambda+o(T^{-4/5})$$
$$=O(T^{-4/5})$$

(b) $$\sup_{\mathscr{X}_T} \operatorname{IVAR}(f_T^{(i)})=2\pi\int_{-c_2}^{c_2} W(\alpha)^2\,d\alpha\,\frac{TH_{4,T}}{H_{2,T}^2}\sup_{\mathscr{X}_T}\int_{-\pi}^{\pi}\frac{N}{T}\,d\lambda+o(T^{-4/5})$$
$$=O(T^{-4/5})$$

(c) $$\sup_{\mathscr{X}_T} \operatorname{IMSE}(f_T^{(i)})=O(T^{-4/5})$$

The same results hold for $\operatorname{SBIAS}(f_T^{(i)})$, $\operatorname{SVAR}(f_T^{(i)})$, *and* $\operatorname{SMSE}(f_T^{(i)})$.

The first equations in (a) and (b) are the same as in classical considerations (with a fixed spectral density). Especially, we obtain the same rate of convergence for the integrated relative mean square error, namely $O(T^{-4/5})$. However, we note that we have made strong use of the local bandwith (properties (3.6) and (3.7)) to obtain these quations.

In the next section we prove that both, the data taper and the local bandwith, are necessary for this uniform rate of convergence.

The factor $\frac{TH_{4,T}}{H_{2,T}^2}$ in Theorem 3.6(b) is larger than 1 (Cauchy-Schwarz inequality). However, it is not correct to conclude from this that tapering always increases the variance of the estimates. We discuss this point in Sect. 4.4. The condition $\kappa<1/40$ allows the choice of a taper with $\lim_{T\to\infty}\frac{TH_{4,T}}{H_{2,T}^2}=1$ (cp. Ex. 3.2).

4. Other Window Estimates

In this section we describe several negative effects that may occur for window estimates if no data taper or no local bandwith is used.

4.1 A Simulation Example

To illutrate the effects we present a simulation example. $T=256$ Gaussian observations were generated for an AR(14)-process with innovation variance 1 and characteristic roots $z_j = q_j e^{i\lambda_j}$ and $\bar{z}_j$ where

$$\begin{array}{ll} q_1=0.95 & \lambda_1=0.5 \\ q_2=0.95 & \lambda_2=1.0 \\ q_3=0.99 & \lambda_3=1.5 \\ q_4=0.99 & \lambda_4=1.5 \\ q_5=0.95 & \lambda_5=2.0 \\ q_6=0.95 & \lambda_6=2.5 \\ q_7=0.95 & \lambda_7=2.5 \end{array}$$

(The same process was used in (Dahlhaus, 1988) for the consideration of nonparametric estimates). Afterwards different estimates of the spectral density (based on the same realization of the process) were considered. In all figures we have plotted the log spectrum and the log spectral estimate.

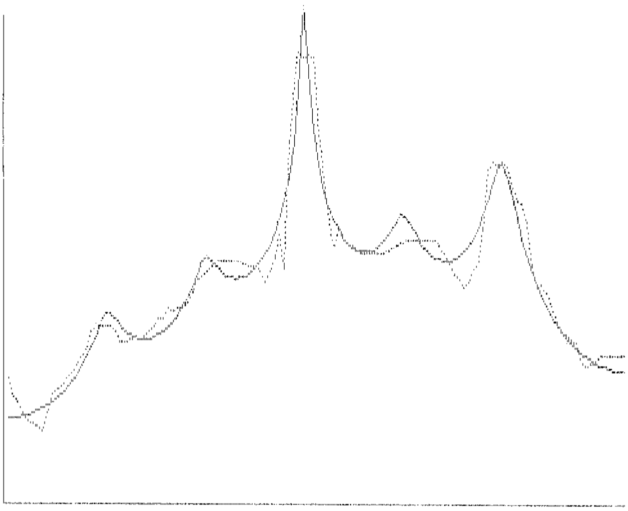

Fig. 2. High resolution spectral estimate

In Fig. 2 we see the true spectral density and the high resolution estimate of Sect. 3 with the Tukey-Hamming taper

$$h_{t,T} = \tfrac{1}{2}\,[1-\cos\{2\pi(t+0.5)/T\}],$$

the Priestley window

$$W(\lambda) = \frac{3}{4\pi}\left(1-\left(\frac{\lambda}{\pi}\right)^2\right) \qquad |\lambda| \geqq \pi,$$

and a local bandwith of the form (3.5) where $(\log f)''$ and $(\log f)'$ were estimated from a preliminary window estimate with a global bandwith. Although, the estimate is not bad we feel that some improvements, in particular concerning the bandwith selection, could be made. For example, the small sidelobs beside the strong peak are disturbing.

4.2 Global Bandwith Selection

With a global bandwith selection we obtain only a lower rate of convergence. The reason is that we need a very small bandwith B_T (a large N) at the peaks which prevents a good rate of convergence. This is made precise in the next theorem.

(4.1) Theorem. *Let $s_1, s_2 \in \mathbb{N}$, $s_1+s_2 \geqq 1$, $\delta_0>0$, $C_0>1$, and $(h_{t,T})_{T\in\mathbb{N}}$ be a sequence of data tapers of degree (k,κ) with $k \geqq \max\{s_1, s_2, s_1+s_2-1\}$ and $\kappa<1/40$. Furthermore, let $N=N(T)$ be independent of f and λ with $N\to\infty$.*

(a) *If in addition $s_1=0$, $NT^{2\kappa-1}\log^2 T\to 0$, $\delta_0<\frac{\pi}{3}$ and $C_0 \geqq 20$, then there exists a constant $C>0$ with*

$$(a.1) \quad \sup_{\mathscr{X}_T} \int_{-\pi}^{\pi} \left(\frac{Ef_T^{(i)}(\lambda)}{f(\lambda)}-1\right) d\lambda \geqq C\,\frac{T^{2s_2-1}}{N^{2s_2}}$$

$$(a.2) \quad \sup_{\mathscr{X}_T} \mathrm{IBIAS}\,(f_T^{(i)}) \geqq C\,\frac{T^{4s_2-2}}{N^{4s_2}}$$

$$(a.3) \quad \sup_{\mathscr{X}_T} \mathrm{IMSE}\,(f_T^{(i)}) \geqq C T^{-2/(4s_2+1)}$$

for $i=1,2$ and all $T \geqq T_0$ with some $T_0\in\mathbb{N}$. If $s_2=0$ the same holds with s_2 replaced by s_1.

(b) *If in addition*

$$N \text{ is independent of } f \text{ and } \lambda \text{ with } N \geqq T^{(4s_1+4s_2-2.5)/(4s_1+4s_2-2)} \text{ and } N/T\to 0, \quad (4.1)$$

then we have for $i=1,2$

$$(b.1) \quad \sup_{\mathscr{X}_T} \mathrm{IBIAS}\,(f_T^{(i)}) = o(1)$$

$$(b.2) \quad \sup_{\mathscr{X}_T} \mathrm{IVAR}\,(f_T^{(i)}) = o(1)$$

$$(b.3) \quad \sup_{\mathscr{X}_T} \mathrm{IMSE}\,(f_T^{(i)}) = o(1)$$

All results of (a) and (b) also hold for the sum statistics SBIAS, SVAR and SMSE.

(4.2) *Remark.* As a consequence of Theorem 4.1 we see that the rate of convergence of $\sup_{\mathscr{X}_T} \mathrm{IMSE}\,(f_T^{(i)})$ with a global bandwith is not as good as the rate with a suitable local bandwith. Furthermore, the rate decreases with the multiplicity of the peak s_2. The same holds if troughs are present ($s_1>0$). We are convinced that it is even possible to prove $\sup_{\mathscr{X}_T} \mathrm{IBIAS}\,(f_T^{(i)}) \approx \frac{T^{4s_2-1}}{N^{4s_2}}$. However, this would require much more calculations.

Figure 1 in the introduction shows the same estimate as in Fig. 2 but with a global bandwith (dashed line). We see that the sharp peak is too broad.

4.3 The Leakage and the Trough Effect

We now study the behaviour of the estimate if no data taper is used. In this section we discuss the bias of the estimate and prove that the window estimate with the nontapered periodogram may even be inconsistent if the spectrum contains strong peaks (leakage effect) or strong troughs. More generally, we consider the tapered periodogram where the degree of the taper is too low.

(4.3) Theorem. *Let* $(s_1 \leqq 1, s_2 > 0)$ *or* $(s_1 > 0, s_2 \leqq 1)$. *Suppose that the applied data taper is of degree* (k, κ) *with* $k < \max\{s_1, s_2, s_1 + s_2 - 1\}$ *and* $\kappa < 1/10$, *and N fulfills* (3.6) *and* (3.7) *or* (4.1). *Then we have for* $C_0 \geqq 20$ *and* $0 < \delta_0 < \pi/3$ *with a constant* $C > 0$

$$\text{(a)}\quad \sup_{\mathscr{X}_T} \int_{-\pi}^{\pi} \left(\frac{E f_T^{(i)}(\lambda)}{f(\lambda)} - 1 \right) d\lambda \geqq C,$$

$$\text{(b)}\quad \sup_{\mathscr{X}_T} \operatorname{IBIAS}(f_T^{(i)}) \geqq C$$

and

$$\text{(c)}\quad \sup_{\mathscr{X}_T} \operatorname{IMSE}(f_T^{(i)}) \geqq C$$

for $i = 1, 2$ *and all* $T \geqq T_0$ *with some* $T_0 \in \mathbb{N}$. *The same holds for the sum statistics* SBIAS *and* SMSE.

Theorem 4.3 together with Theorem 3.6 proves that a data taper of degree (k, κ) with $k = \max\{s_1, s_2, s_1 + s_2 - 1\}$ is necessary and sufficient for the window estimate to have the high resolution property. If the degree is not sufficient we do not even have consistency of the estimate in the above sense. Thus we need a certain smoothness of the data taper at the ends of the observation domain. In the nontapered case we obtain the following result.

(4.4) Corollary. *Let* $s_1, s_2 \in \mathbb{N}_0$, $s_1 + s_2 \geqq 1$, $C_0 \geqq 20$ *and* $0 < \delta_0 < \pi/3$. *Suppose that no data taper is applied and N fulfills* (3.6) *and* (3.7) *or* (4.1). *Then all assertions of Theorem 4.3 hold.*

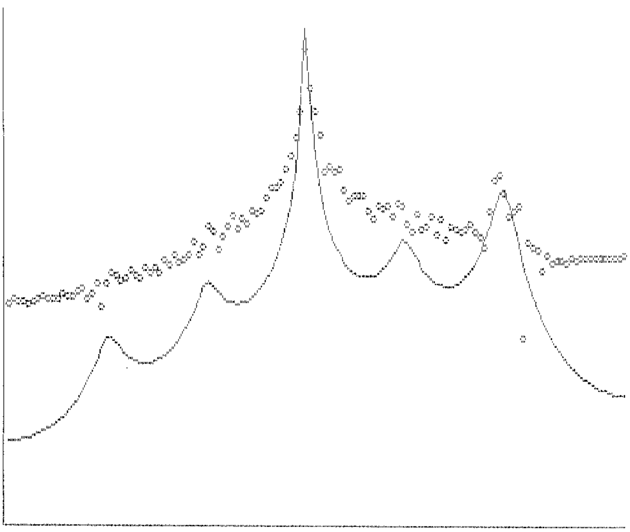

Fig. 3. Periodogram without data taper

The corollary establishes theoretically the leakage effect for window estimates with the nontapered periodogram ($s_2 > 0$). The spectrum is overestimated due to leakage from strong peaks. From Theorem 3.6 we see that this effect can be cured by applying a data taper.

In Fig. 3 we see the periodogram without a data taper together with the true spectral density. The corresponding window estimate with a global bandwith is plotted in Fig. 1 (dotted line). The spectrum is overestimated and obviously the same will hold for any kernel estimate.

Corollary 4.4 also establishes theoretically the trough effect ($s_1 > 0$). Again the spectrum is *over*estimated. It is not possible to find the troughs sufficiently with the nontapered periodogram. This effect can be cured by applying a data taper.

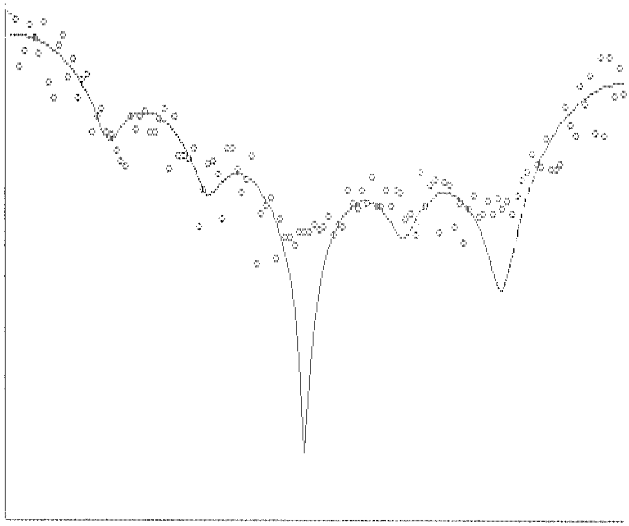

Fig. 4. Periodogram without data taper

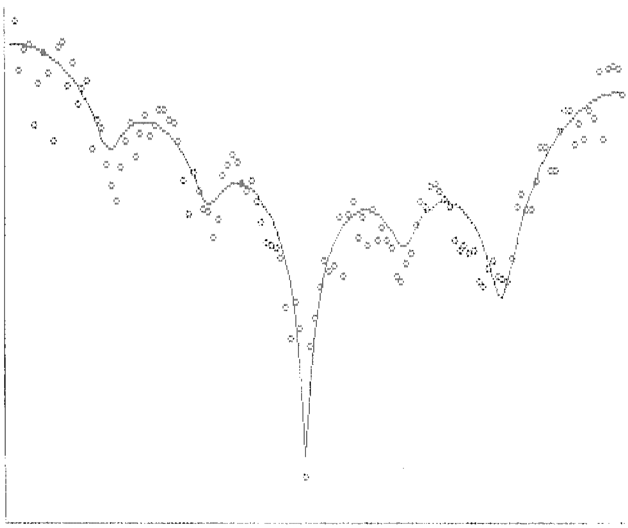

Fig. 5. Periodogram with data taper

In Figs. 4 and 5 we see an example for the trough effect. Instead of an AR(14)-process we have simulated a MA(14)-process with the same roots. In Fig. 4 we have plotted the nontapered periodogram (again unsmoothed) and in Fig. 5 the tapered periodogram. The nontapered estimate is not able to resolve the troughs, while the tapered estimate clearly is. In the nontapered case the spectrum is overestimated. This effect is of importance if one takes differences to remove trends or seasonal differences to remove periodic components, and one wants to decide with a nonparametric estimate whether the difference filter was too strong.

4.4 The Variance Effect

It is a common opinion (cf. Brillinger, 1981, p. 151; Hannan, 1970, p. 272; Priestley, 1981, p. 562) that tapering may reduce in many situations the bias while it increases the variance. This opinion results from the term $\dfrac{TH_{4,T}}{H_{2,T}^2}$ in the asymptotic variance of $f_T^{(i)}$ (in this paper Theorem 3.6(b)) which is greater than one in the tapered case and equal to one in the nontapered case. However, this argument implicitely assumes that the variance converges, in the situation where strong peaks are present and no data taper is applied, to the same limit as in Theorem 3.6(b) with $TH_{4,T}/H_{2,T}^2=1$. In the next theorem we prove that this is not true.

(4.5) Theorem. *Let* $s_1>0$ *or* $s_2>0$. *Suppose that no data taper is applied and* N *fulfills* (3.6) *and* (3.7) *or* (4.1). *Then we have for* $C_0 \geqq 20$ *and* $0<\delta_0<\pi/3$

$$\sup_{\mathscr{X}_T} \mathrm{IVAR}(f_T^{(i)}) \geqq C$$

for $i=1,2$ *with some* $C>0$. *The same holds for the sum statistics* SVAR.

We conjecture that the same assertion as in Theorem 4.5 holds if the degree of the applied data taper is too low, i.e. if $k<\max\{s_1, s_2, s_1+s_2-1\}$.

5. Segment Estimates

We now study estimates of the spectral density obtained by averaging periodograms over (overlapping) data segments. Let

$$f_T^{(3)}(\lambda) := \frac{1}{M} \sum_{k=0}^{M-1} I_N^{Lk}(\lambda)$$

with

$$I_N^{Lk}(\lambda) = \{2\pi H_{2,N}(0)\}^{-1} \left| \sum_{t=Lk}^{Lk+N-1} h_{t-Lk,N} X_t \exp\{-i\lambda(t-Lk)\} \right|^2 .$$

The interesting cases are $L=N$ and $L<N$ where the segments are overlapping. $T=LM+N-L$ is the sample size. We shall call this estimate the segment estimate for short. The estimate has been considered by various authors (e.g. Bartlett, 1950; Welch, 1967; Brillinger, 1975; Kolmogorov and Zhurbenko, 1978; Dahlhaus, 1985). For smooth spectral densities this estimate has roughly the same mean

sequare error as the window estimate (with global bandwith). Zhurbenko (1980) proves for a particular taper $h_{t,N}$ and Lipschitz-continuous spectral densities that the mean square error of the estimate is lower than the mean square error of the window estimate with several common windows but higher than the optimal window estimate. Furthermore, Zhurbenko (1983) shows by considering a spectral measure with a jump that the segment estimate is less sensitive to disturbances from outlying frequencies than the window estimate (Zhurbenko, 1983, Theorem 8 and Theorem 11). However, he compares the segment estimate with taper to the window estimate without taper.

We now study the segment estimate in the framework of this paper and prove that the uniform rate of convergence over the class $\mathscr{X}_T$ is lower than the corresponding convergence rate of the window estimate. We do not make any assumptions on the relation between L and N.

(5.1) Theorem. *Let* $s_1, s_2 \in \mathbb{N}$ *with* $s_1+s_2 \geqq 1$, $C_0 \geqq 20$, $0<\delta_0<\frac{\pi}{3}$, *and* C *be a positive constant.*

(a) *If no taper is applied, i.e.* $h_{t,N}=1$ $(t=0,\ldots,N-1)$, *then*

$$\sup_{\mathscr{X}_T} \operatorname{IVAR}(f_T^{(3)}) \geqq C\,\frac{T^2}{N^2} \geqq C.$$

(b) *If the data taper is of degree* (k,κ) *with* $k \geqq 1$ *and* $N^{1+8\kappa}\log N \log T/T \to 0$, *then*

$$\begin{aligned}\sup_{\mathscr{X}_T} \operatorname{IVAR}(f_T^{(3)}) &\geqq C T^{-(1+4\kappa)/(4+4\kappa)}\\ &\geqq C T^{-1/3} \quad \text{if} \quad \kappa<1/8.\end{aligned} \tag{5.1}$$

(c) *If* $h_{t,N}=h_{t,K,P}$ *is the Kolmogorov-Zhurbenko Taper* (cf. *Zhurbenko*, 1980, (2.10)) *with* $K, P \in \mathbb{N}$, $N=K(P-1)$, $K \leqq N^\kappa$ *and* $N^{1+8\kappa}\log N \log T/T \to 0$, *then we also have* (5.1).

The same results hold for the IMSE.

Thus, the IMSE of the segment estimate has a lower uniform rate of convergence than the window estimate of Sect. 3. In fact, the rate is even worse. By considering an AR(s_2)-process in the proof it can be shown that e.g. (5.1) can be replaced by

$$\sup_{\mathscr{X}_T} \operatorname{IVAR}(f_T^{(3)}) \geqq C T^{-(1+4\kappa)/(2s_2+4\kappa)}.$$

By calculating also the bias it is possible to prove

$$\sup_{\mathscr{X}_T} \operatorname{IMSE}(f_T^{(3)}) \geqq C T^{-(1+4\kappa)/(4s_2+1+4\kappa)}.$$

A considerable improvement of the segment estimate may be achieved by using a local segment length N depending on f and the frequency λ, e.g. the segment length N defined by (3.5). However, we doubt that it is possible to achieve by a local segment selection the same rate $T^{-4/5}$ as for the window estimate in Sect. 3. Furthermore, a local segment selection is very inconvenient in practice because the use of the Fast-Fourier algorithm does no longer lead to any computational savings in comparison to an ordinary Fourier transformation. This increases the computational effort dramatically.

6. Concluding Remarks

In this paper we have introduced a mathematical model to describe theoretically nonparametric 'high resolution spectral estimates'. Instead of the ordinary rate of convergence we have studied a uniform rate of convergence over a (with the sample size T) increasing class of stationary processes.

By using this model we were able to prove the advantages of using data tapers in time series analysis. We have explained the leakage effect caused by peaks in the spectrum and the trough effect caused by values close to zero. We could also prove, contrary to widespread conjectures, that tapering does not only reduce the bias but may also reduce the variance of the window estimate.

We also demonstrated the advantages of a local bandwith over a global bandwith. Furthermore, we have proved that segment estimates have a lower uniform rate of convergence.

Several problems remain unsolved. For example, the construction of a bandwith that does not depend on the unknown spectral density but on the data (e.g. on a preliminary estimate) and that leads to a high resolution estimate in the sense of this paper. Furthermore, the paper does not sufficiently answer the question how the data taper should be chosen in practise.

Appendix

The appendix contains the technical details of the paper.

A.1 Properties of the Function $L_T(\alpha)$

(A.1) Lemma. *Let $L_T(\alpha)$ be defined as in* (1.2), $r, s>0$ *and* $\alpha, \beta, \gamma, \nu, \mu \in \mathbb{R}$. *We obtain with a constant K independent of T, T_1 and T_2*

a) $L_T(\alpha)$ *is monotone increasing in T and decreasing in* $\alpha \in [0, \pi]$.

b) $\int_\Pi L_T(\alpha)^r d\alpha \leqq K T^{r-1}$ *for all* $r>1$.

c) $\int_\Pi L_T(\alpha) d\alpha \leqq K \log T$

d) $\pi^{-1} \leqq L_T(\alpha)$

e) $|\alpha| L_T(\alpha) \leqq 1$ *for* $|\alpha| \leqq \pi$

f) $L_{T_1}(\nu)^r L_{T_2}(\mu)^s \leqq L_{T_1}\left(\frac{\nu-\mu}{2}\right)^r L_{T_2}(\mu)^s + L_{T_1}(\nu)^r L_{T_2}\left(\frac{\nu-\mu}{2}\right)^s$ *for* $|\lambda|$, $|\mu| \leqq \pi$

g) $L_T(c\alpha) \leqq K_c L_T(\alpha)$

h) $\int_\Pi L_{T_1}(\alpha+\beta) L_{T_2}(\beta) d\beta \leqq K L_{\min(T_1, T_2)}(\alpha) \max(\log T_1, \log T_2)$

i) $L_T(\alpha) \leqq K L_T(\beta)$ *for all* α, β *with* $|\alpha-\beta| \leqq 2\pi/T$.

Proof. The proofs are straightforward. Some of them may be found in Dahlhaus (1983). To prove f) consider the cases $|\nu| \geqq \frac{|\nu-\mu|}{2}$ and $|\mu| \geqq \frac{|\nu-\mu|}{2}$. To prove h) for $T_1 \leqq T_2$ consider the cases $|\alpha| \leqq T_1^{-1}$ and $|\alpha| \geqq T_1^{-1}$ and apply f) and g).

A.2 The High Resolution Property: Proof of Theorem 2.4

Let $X_t \in \mathscr{X}_T$. Then X_t has a spectral density of the form $f(\lambda)=h_1(\lambda)h_2(\lambda)h_0(\lambda)$ where $h_0(\lambda)=f_{2,Y}(\lambda)$ and

$$h_i(\lambda)=\left\{\prod_{j=1}^{r_i} g_{ij}(\lambda-\lambda_{ij})^{s_{ij}}\right\}^{(-1)^{i+1}} \qquad (i=1,2).$$

We start by proving similar assertions as in Theorem 2.4 for h_1 and h_2 separately. Let $r_1, r_2, m_1, m_2 \in \mathbb{N}$, $T \in \mathbb{N}$, $\underline{\lambda}_1 \in \Pi^{r_1}$, $\underline{\lambda}_2 \in \Pi^{r_2}$, $\underline{T}_1 \in \mathbb{R}_+^{r_1}$, $\underline{T}_2 \in \mathbb{R}_+^{r_2}$. We define for $\ell \in \mathbb{N}_0$

$$R_1(\lambda,\alpha,\ell,m_1)=R_1(\lambda,\alpha,\ell,m_1,r_1,\underline{\lambda}_1,\underline{T}_1)$$
$$=\sum_{j=1}^{r_1}\left[\sum_{k=\ell}^{m_1}|\alpha|^k L_{T_{1j}}(\lambda-\lambda_{1j})^k+\{m_1=2,\ell=3\}|\alpha|^3 L_{T_{1j}}(\lambda-\lambda_{1j})^2\right]$$

$$R_2(\lambda,\alpha,\ell,m_2)=R_2(\lambda,\alpha,\ell,m_2,r_2,\underline{\lambda}_2,\underline{T}_2)$$
$$=\sum_{j_2,j_3=1}^{r_2}\left[\sum_{\substack{k_2=1,\dots,m_2\\ k_3=0,\dots,\max(0,\min(\ell-1,m_2))\\ \ell\le k_2+k_3}}|\alpha|^{k_2+k_3}L_{T_{2j_2}}(\lambda+\alpha-\lambda_{2j_2})^{k_2}L_{T_{2j_3}}(\lambda-\lambda_{2j_3})^{k_3}\right]$$

(A.2) Lemma. *We have for $\ell=1,2,3$ and $i=1,2$*

$$\frac{h_i(\lambda+\alpha)}{h_i(\lambda)}-1=\sum_{k=1}^{\ell-1}\frac{h_i^{(k)}(\lambda)}{h_i(\lambda)}\frac{\alpha^k}{k!}+O[R_i(\lambda,\alpha,\ell,2s_i)].$$

The O-term depends only on s_1, s_2, C_0 and δ_0.

Proof. Let $i=1$. Then

$$\frac{g_{1j}(\lambda+\alpha)}{g_{1j}(\lambda)}-1=\alpha\frac{g'_{1j}(\lambda)}{g_{1j}(\lambda)}+\frac{\alpha^2}{2}\frac{g''_{1j}(\lambda)}{g_{1j}(\lambda)}+O[|\alpha|^3L_{T_{1j}}(\lambda)^2].$$

The relations

$$\left(\prod_{j=1}^{n}x_j\right)-1=\sum_{\substack{M\subset\{1,\dots,n\}\\ M\neq\phi}}\prod_{j\in M}(x_j-1), \tag{A.1}$$

(2.3) and (2.4) imply

$$\frac{g_{1j}(\lambda+\alpha)^{s_{1j}}}{g_{1j}(\lambda)^{s_{1j}}}-1=\alpha\frac{(g_{1j}(\lambda)^{s_{1j}})'}{g_{1j}(\lambda)^{s_{1j}}}+\frac{\alpha^2}{2}\frac{(g_{1j}(\lambda)^{s_{1j}})''}{g_{1j}(\lambda)^{s_{1j}}}$$
$$+O\left[\sum_{k=3}^{2s_{1j}}|\alpha|^kL_{T_{1j}}(\lambda)^k+\{s_{1j}=1\}|\alpha|^3L_{T_{1j}}(\lambda)^2\right] \tag{A.2}$$

Lemma A.1(f) now implies with (A.1) the result for $\ell=3$ and with (2.3) and (2.4) also for $\ell=1,2$.

The case $i=2$ is more difficult. Let $h_3 = h_2^{-1}$. We obtain as for $i=1$

$$\frac{h_2(\lambda+\alpha)}{h_2(\lambda)}-1=\frac{h_3(\lambda)}{h_3(\lambda+\alpha)}-1=(-\alpha)\frac{h_3'(\lambda+\alpha)}{h_3(\lambda+\alpha)}+\frac{\alpha^2}{2}\frac{h_3''(\lambda+\alpha)}{h_3(\lambda+\alpha)}$$
$$+O\left[\sum_{j=1}^{r_2}\sum_{k=3}^{2s_2}|\alpha|^k L_{T_{2j}}(\lambda+\alpha-\lambda_{2j})^k+\{s_2=1,\ell=3\}|\alpha|^3 L_{T_{2j}}(\lambda+\alpha-\lambda_{2j})^2\right].$$

This implies the result for $\ell=1$. We now replace $\frac{h_3'(\lambda+\alpha)}{h_3(\lambda+\alpha)}$ and $\frac{h_3''(\lambda+\alpha)}{h_3(\lambda+\alpha)}$ by functions depending on λ. Below we will prove

$$\alpha\left(\frac{h_3'(\lambda+\alpha)}{h_3(\lambda+\alpha)}-\frac{h_3'(\lambda)}{h_3(\lambda)}\right)=O[R_2(\lambda,\alpha,2,2s_2)] \qquad \text{(A.3)}$$

which implies the result for $\ell=2$. Similarly, it can be shown

$$\alpha^2\left(\frac{h_3''(\lambda+\alpha)}{h_3(\lambda+\alpha)}-\frac{h_3''(\lambda)}{h_3(\lambda)}\right)=O[R_2(\lambda,\alpha,3,2s_2)] \qquad \text{(A.4)}$$

and

$$\alpha\left(\frac{h_3'(\lambda+\alpha)}{h_3(\lambda+\alpha)}-\frac{h_3'(\lambda)}{h_3(\lambda)}\right)=\alpha^2\left[\frac{h_3''(\lambda)}{h_3(\lambda)}-\left(\frac{h_3'(\lambda)}{h_3(\lambda)}\right)^2\right]+O[R_2(\lambda,\alpha,3,2s_2)]. \qquad \text{(A.5)}$$

Since $\frac{h_2'}{h_2}=-\frac{h_3'}{h_3}$ and $\frac{h_2''}{h_2}=-\frac{h_3''}{h_3}+2\left(\frac{h_3'}{h_3}\right)^2$ this implies the result for $\ell=3$.

We have

$$\frac{h_3'(\lambda+\alpha)}{h_3(\lambda+\alpha)}-\frac{h_3'(\lambda)}{h_3(\lambda)}=\sum_{j=1}^{r_2}s_{2j}\left[\frac{g_{2j}'(\lambda+\alpha)}{g_{2j}(\lambda+\alpha)}-\frac{g_{2j}'(\lambda)}{g_{2j}(\lambda)}\right].$$

Since

$$\alpha\left(\frac{g_{2j}'(\lambda+\alpha)}{g_{2j}(\lambda+\alpha)}-\frac{g_{2j}'(\lambda)}{g_{2j}(\lambda)}\right)=\alpha\,\frac{g_{2j}'(\lambda+\alpha)-g_{2j}'(\lambda)}{g_{2j}(\lambda+\alpha)}+\alpha\,\frac{g_{2j}'(\lambda)}{g_{2j}(\lambda)}\left[\frac{g_{2j}(\lambda)}{g_{2j}(\lambda+\alpha)}-1\right]$$
$$=O[|\alpha|^2 L_{T_{2j}}(\lambda+\alpha-\lambda_{2j})^2+|\alpha|^2 L_{T_{2j}}(\lambda-\lambda_{2j})L_{T_{2j}}(\lambda+\alpha-\lambda_{2j})$$
$$+|\alpha|^3 L_{T_{2j}}(\lambda-\lambda_{2j})L_{T_{2j}}(\lambda+\alpha-\lambda_{2j})^2] \qquad \text{(A.6)}$$

we obtain (A.3).

Proof of Theorem 2.4. (a) The result follows with Lemma A.2 and (A.1).

(b) Let $x=|\alpha|L_{T_{1j}}(\lambda)$ and $y=\sum_{k=1}^{s_{1j}}x^k$. In (A.2) we have proved that

$$\left|\frac{g_{1j}(\lambda+\alpha)^{s_{1j}}}{g_{1j}(\lambda)^{s_{1j}}}-1\right|\leqq K\sum_{k=1}^{2s_{1j}}x^k\leqq Ky+(Ky)^2.$$

This implies

$$\left|\left|\frac{A_{1j}(\lambda+\alpha)}{A_{1j}(\lambda)}\right|^{s_{1j}}-1\right|\leqq 2Ky.$$

In the same way we get

$$\left\| \frac{A_{2j}(\lambda)}{A_{2j}(\lambda+\alpha)} \right|^{s_{2j}} - 1 \Bigg| \leqq 2K \sum_{k=1}^{s_{2j}} |\alpha|^k L_{T_{2j}}(\lambda+\alpha)^k$$

and we therefore obtain with Lemma A.1(f) and (A.1) the result. Since (b) implies

$$\left| \frac{A(\lambda+\alpha)}{A(\lambda)} \right| = O[R(\lambda, \alpha, 0, s_1, s_2)]$$

we also obtain (c).

A.3 Local Bandwith Selection and Proof of Lemma 3.4

(A.3) Lemma. *Let* W: $\mathbb{R} \to [0, c_1] (c_1 > 1)$ *with* $W(\beta) = 0$ *for* $|\beta| > c_2$, $\int_{-c_2}^{c_2} W(\beta) d\beta = 1$, *and* $W_N(\alpha) = N W(N\alpha)$. *Then*

(a) $$W_N(\alpha) \leqq c_1^{k+1} N^{-k+1} L_N(\alpha)^k \quad \textit{for all} \quad k \in \mathbb{N}.$$

If in addition $T^{1/5} L_{T_1}(\lambda)^{4/5} \leqq N \leqq T$ *with* $T_1 \leqq T$, *then*

(b) $$\int_{-\pi}^{\pi} |\alpha|^k L_{T_1}(\lambda+\alpha)^\ell W_N(\alpha) d\alpha \leqq K \frac{L_{T_1}(\lambda)^\ell}{N^k} \log N$$

and

(c) $$\int_{-\pi}^{\pi} |\alpha|^k L_{T_1}(\lambda+\alpha)^\ell L_T(\mu+\alpha)^2 W_N(\alpha) d\alpha \leqq K_p N^{1-1/p} T^{1+1/p} \frac{L_{T_1}(\lambda)^\ell}{N^k} \log N$$

for $p > 1$ *and all* $k, \ell \in \mathbb{N}_0$ *with constants that are independent of* N, T and λ.

Proof. (a) is straightforward. To prove (b) we start with $T_1 \leqq N$. We obtain with Lemma A.1(i)

$$\int_{-\pi}^{\pi} |\alpha|^k L_{T_1}(\lambda+\alpha)^\ell W_N(\alpha) d\alpha = \int_{-c_2/N}^{c_2/N} |\alpha|^k L_{T_1}(\lambda+\alpha)^\ell W_N(\alpha) d\alpha$$
$$\leqq K L_{T_1}(\lambda)^\ell \int_{-\pi}^{\pi} |\alpha|^k W_N(\alpha) d\alpha \leqq K \frac{L_{T_1}(\lambda)^\ell}{N^k}.$$

If $N \leqq T_1 \leqq T$ we consider the three cases $|\lambda| \leqq 2c_2/T_1$, $2c_2/T_1 \leqq |\lambda| \leqq 2c_2/N$ and $|\lambda| \geqq 2c_2/N$ separately. We omit details.

(c) We apply Hölder's inequality and obtain as an upper bound

$$K \left\{ \int_{-\pi}^{\pi} |\alpha|^{kp} L_{T_1}(\lambda+\alpha)^{\ell p} W_N(\alpha) d\alpha \right\}^{1/p} \left\{ \int_{-\pi}^{\pi} L_T(\mu+\alpha)^{\frac{2p}{p-1}} W_N(\alpha) da \right\}^{1-1/p}.$$

The second terms is bounded by

$$K N^{1-1/p} T^{1+1/p}$$

which implies with (b) the result.

Proof of Lemma 3.4. We have

$$(\log f(\lambda))' = \sum_{i=1}^{2} (-1)^{i+1} \sum_{j=1}^{r_i} s_{ij} \frac{g'_{ij}(\lambda-\lambda_{ij})}{g_{ij}(\lambda-\lambda_{ij})} + \frac{f'_Y(\lambda)}{f_Y(\lambda)} \tag{A.7}$$

and

$$(\log f(\lambda))'' = \sum_{i=1}^{2} (-1)^{i+1} \sum_{j=1}^{r_i} s_{ij} \left[\frac{g''_{ij}(\lambda-\lambda_{ij})}{g_{ij}(\lambda-\lambda_{ij})} - \left(\frac{g'_{ij}(\lambda-\lambda_{ij})}{g_{ij}(\lambda-\lambda_{ij})}\right)^2\right] + \frac{f''_Y(\lambda)}{f_Y(\lambda)} - \left(\frac{f'_Y(\lambda)}{f_Y(\lambda)}\right)^2$$

(2.2) to (2.4) and Lemma A.1(f) now imply (3.6).

Before proving (3.7) we make a preliminary consideration. If $|\lambda-\lambda_{ij}| \geqq \delta_0$ for all i,j then (2.2) to (2.4) and Lemma A.1(e) imply $\left(\text{note that } |r_1|, |r_2| \leqq \frac{\pi}{\delta_0}\right)$

$$|(\log f(\lambda))'| \leqq (s_1+s_2)\, \frac{\pi C_0^2}{\delta_0^2} + C_0^2 \tag{A.8}$$

and

$$|(\log f(\lambda))''| \leqq (s_1+s_2)\, \frac{2\pi C_0^4}{\delta_0^3} + 2C_0^4. \tag{A.9}$$

Let

$$K_0 = \max\left[(s_1+s_2)\, \frac{2\pi C_0^4}{\delta_0^3} + 2C_0^4,\ (s_1+s_2)\, \frac{\pi C_0^2}{\delta_0^2} + C_0^2,\ \frac{1}{8C_0^4\delta_0}\right]$$

and $C_2 = \{64 C_0^8 K_0^2\}^{-1}$. To prove (3.7) we now consider three cases. We start with $|\lambda-\lambda_{ij}| \leqq \delta := \frac{1}{8C_0^4 K_0}$ and $T_{ij} \geqq 8C_0^4 K_0$ for some i,j. Since $K_0 \geqq \frac{1}{8C_0^4 K_0}$ we have $\delta \leqq \delta_0$ and therefore $|\lambda-\lambda_{k\ell}| > \delta_0$ for $(k,\ell) \neq (i,j)$. Thus,

$$C_2 \max_{k,\ell} L_{T_{k\ell}}(\lambda-\lambda_{k\ell})^2 = C_2 L_{T_{ij}}(\lambda-\lambda_{ij})^2.$$

Furthermore, we have

$$(\log f(\lambda))' = (-1)^{i+1} s_{ij}\{y(\lambda) + R_{ij}^{(1)}(\lambda)\}$$

and

$$(\log f(\lambda))'' = (-1)^{i+1} s_{ij}\{x(\lambda) - y(\lambda)^2 + R_{ij}^{(2)}(\lambda)\}$$

with $x(\lambda) = g''_{ij}(\lambda-\lambda_{ij})/g_{ij}(\lambda-\lambda_{ij})$, $y(\lambda) = g'_{ij}(x-\lambda_{ij})/g_{ij}(\lambda-\lambda_{ij})$. Analogously to (A.8) and (A.9) we obtain for the remainder terms $|R_{ij}^{(1)}(\lambda)| \leqq K_0$, and $|R_{ij}^{(2)}(\lambda)| \leqq K_0$. Elementary calculations give

$$|(\log f(\lambda))''| + (\log f(\lambda))'^2 \geqq x(\lambda) + 2y(\lambda) R_{ij}^{(1)}(\lambda) + R_{ij}^{(1)}(\lambda)^2 + R_{ij}^{(2)}(\lambda). \tag{A.10}$$

(2.2) to (2.5) imply

$$|y(\lambda)|^2 \leqq C_0^4 L_{T_{ij}}(\lambda-\lambda_{ij})^2 \leqq C_0^6 |x(\lambda)|$$

and, since $\delta \geqq T_{ij}^{-1}$,

$$|x(\lambda)| \geqq 64 C_0^6 K_0^2.$$

This gives

$$2|y(\lambda)R_{ij}^{(1)}(\lambda)| \leqq \frac{|x(\lambda)|}{4}, \quad |R_{ij}^{(1)}(\lambda)|^2 \leqq \frac{|x(\lambda)|}{4}, \text{ and } |R_{ij}^{(2)}(\lambda)| \leqq \frac{|x(\lambda)|}{4},$$

i.e. (A.10) is larger than $\frac{|x(\lambda)|}{4} \geqq C_2 L_{T_{ij}}(\lambda-\lambda_{ij})^2$.

If $|\lambda-\lambda_{ij}| \leqq \delta = \frac{1}{8C_0^4 K_0}$ and $T_{ij} \leqq 8C_0^4 K_0$ we have

$$1 \geqq C_2 T_{ij}^2 \geqq C_2 L_{T_{ij}}(\lambda-\lambda_{ij})^2$$

which implies (3.7). If $|\lambda-\lambda_{k\ell}| \geqq \delta$ for all k,ℓ we obtain

$$C_2 \max_{k,\ell} L_{T_{k\ell}}(\lambda-\lambda_{k,\ell})^2 \leqq C_2 \frac{1}{\delta^2} = 1$$

which again implies (3.7).

A.4 High Resolution Window Estimates

Proof of Theorem 3.6. Let $i=1$. (a) We have with $K_T(\alpha) = \{2\pi H_{2,T}\}^{-1} |H_T(\alpha)|^2$

$$\frac{Ef_T^{(1)}(\lambda)}{f(\lambda)} - 1 = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left[\frac{f(\lambda+\alpha+\beta)}{f(\lambda)} - 1\right] W_N(\beta) K_T(\alpha)\, d\alpha\, d\beta$$

$$= \int_{-\pi}^{\pi} \left[\frac{f(\lambda+\beta)}{f(\lambda)} - 1\right] W_N(\beta) \int_{-\pi}^{\pi} \left[\frac{f(\lambda+\alpha+\beta)}{f(\lambda+\beta)} - 1\right] K_T(\alpha)\, d\alpha\, d\beta \tag{A.11}$$

$$+ \int_{-\pi}^{\pi} W_N(\beta) \int_{-\pi}^{\pi} \left[\frac{f(\lambda+\alpha+\beta)}{f(\lambda+\beta)} - 1\right] K_T(\alpha)\, d\alpha\, d\beta \tag{A.12}$$

$$+ \int_{-\pi}^{\pi} \left[\frac{f(\lambda+\beta)}{f(\lambda)} - 1\right] W_N(\beta)\, d\beta \tag{A.13}$$

We obtain with Theorem 2.4(a)($\ell=1$), Lemma 3.3 and Lemma A.1

$$\int_{-\pi}^{\pi} \left[\frac{f(\lambda+\alpha+\beta)}{f(\lambda+\beta)} - 1\right] K_T(\alpha)\, d\alpha$$

$$\leqq K \sum_{\substack{j_i=1 \\ i=1,2}}^{r_i} \sum_{\substack{k_i=0 \\ k_1+k_2 \geqq 1}}^{2s_i} \int_{-\pi}^{\pi} T^{-(2k+1-2\kappa)} L_{T_{1j_1}}(\lambda+\beta-\lambda_{1j_1})^{k_1}$$

$$\cdot L_{T_{2j_2}}(\lambda+\beta+\alpha-\lambda_{2j_2})^{k_2} L_T(\alpha)^{2k+2-k_1-k_2} d\alpha$$

$$\leqq KT^{2\kappa-1} \sum_{j_i} \int_{-\pi}^{\pi} \{L_{T_{1j_1}}(\lambda+\beta-\lambda_{1j_1}) L_{T_{2j_2}}(\lambda+\beta+\alpha-\lambda_{2j_2})$$

$$+ L_T(\alpha)\,[L_{T_{1j_1}}(\lambda+\beta-\lambda_{1j_1}) + L_{T_{2j_2}}(\lambda+\beta+\alpha-\lambda_{2j_2})]\}\, d\alpha$$

$$\leqq KT^{2\kappa-1} \log T \sum_{j_i} \{L_{T_{1j_1}}(\lambda+\beta-\lambda_{1j_1}) + L_{T_{2j_2}}(\lambda+\beta-\lambda_{2j_2})\} \tag{A.14}$$

Using Lemma A.3 we therefore obtain as an upper bound for (A.12)

$$KT^{2\kappa-1}\log^2 T\sum_{j_i}\{L_T(\lambda-\lambda_{1j})+L_T(\lambda-\lambda_{2j})\}.$$

Thus, the integrated square of (A.12) is with Lemma A.1(f) less than

$$KT^{4\kappa-2}\log^4 T\sum_{j_i}\int\{L_T(\lambda-\lambda_{1j})^2+L_T(\lambda-\lambda_{2j})^2\}d\lambda$$

$$\leqq KT^{4\kappa-1}\log^4 T=o(T^{-4/5}).$$

With (A.14), Theorem 2.4(a) ($\ell=1$) and Lemma A.1(f) we obtain as an upper bound of (A.11)

$$KT^{2\kappa-1}\log T\sum_{j_i}\sum_{\substack{k_i=0\\k_1+k_2\geqq 1}}^{2s_i}\int_{-\pi}^{\pi}|\beta|^{k_1+k_2}L_{T_{1j_1}}(\lambda-\lambda_{1j_1})^{k_1}$$

$$\cdot L_{T_{2j_2}}(\lambda+\beta-\lambda_{2j})^{k_2}\{L_{T_{1j_3}}(\lambda+\beta-\lambda_{1j_3})+L_{T_{2j_2}}(\lambda+\beta-\lambda_{2j_2})\}W_N(\beta)d\beta.$$

Lemma A.3(b), the Cauchy-Schwarz inequality, and Lemma A.1(f) lead to the upper bound

$$KT^{2\kappa-1}(\log^2 T)\sum_{j_i}\sum_{k_1+k_2\geqq 2}\frac{L_{T_{1j_1}}(\lambda-\lambda_{1j_1})^{k_1}L_{T_{2j_2}}(\lambda-\lambda_{2j_2})^{k_2}}{N^{k_1+k_2-1}}$$

From (3.7) we have

$$\frac{L_{T_{1j}}(\lambda-\lambda_{ij})^k}{N^k}\leqq K\frac{L_{T_{ij}}(\lambda-\lambda_{ij})^{k/5}}{T^{k/5}}\leqq K. \tag{A.15}$$

Thus, the above expression is bounded by

$$KT^{2\kappa-1}(\log^2 T)\sum_{j_i}\{L_{T_{1j_1}}(\lambda-\lambda_{1j_1})+L_{T_{2j_2}}(\lambda-\lambda_{2j_2})\}.$$

Therefore, we obtain that the integrated square of (A.11) is $o(T^{-4/5})$. Theorem 2.4(a) ($\ell=3$) implies that (A.13) is equal to

$$\frac{1}{2N^2}\frac{f''(\lambda)}{f(\lambda)}\int\alpha^2 W(\alpha)d\alpha+O\left[\int_{-\pi}^{\pi}R(\lambda,\alpha,3,2s_1,2s_2)W_N(\alpha)d\alpha\right]. \tag{A.16}$$

The same arguments as before imply that the remainder is bounded by

$$K\log T\sum_{j_i}\sum_{\substack{k_1=0,\ldots,3\\k_2=0,\ldots,3\\k_1+k_2\geqq 3}}\frac{L_{T_{1j_1}}(\lambda-\lambda_{1j_1})^{k_1/5}L_{T_{2j_2}}(\lambda-\lambda_{2j_2})^{k_2/5}}{T^{(k_1+k_2)/5}}.$$

The integrand square of this expression is $o(T^{-4/5})$. Using Lemma 3.4 the first term in (A.16) is bounded by

$$KT^{-2/5}\sum_{ij}L_{T_{ij}}(\lambda-\lambda_{ij})^{2/5}$$

whose integrated square is $O(T^{-4/5})$. Since all constants depend only on s_1, s_2, δ_0 and C_0 part (a) is proved.

(b) Let

$$\psi_T(\alpha,\gamma)=H_T(\alpha_1-\gamma_1)H_T(\gamma_1-\alpha_2)H_T(\alpha_2-\gamma_2)H_T(\gamma_2-\alpha_1).$$

By using the product theorem for cumulants (Brillinger, 1981, Theor. 2.3.2) we obtain

$$\int_{-\pi}^{\pi} \operatorname{var}\left(\frac{\hat{f}_T^{(1)}(\lambda)}{f(\lambda)}\right) d\lambda=\{2\pi H_{2,T}\}^{-2}$$

$$\cdot \int_{-\pi}^{\pi}\left[\int_{\Pi^4} \frac{f(\gamma_1)f(\gamma_2)}{f(\lambda)f(\lambda)} W_N(\lambda-\alpha_1)W_N(\lambda-\alpha_2)\psi_T(\alpha,\gamma)d\gamma d\alpha\right.$$

$$+\int_{\Pi^4} \frac{f(\gamma_1)f(\gamma_2)}{f(\lambda)f(\lambda)} W_N(\lambda-\alpha_1)W_N(\lambda+\alpha_2)\psi_T(\alpha,\gamma)d\gamma d\alpha$$

$$+\int_{\Pi^5} \frac{f_4(\gamma_1,\gamma_2,\gamma_3)}{f(\lambda)f(\lambda)} W_N(\lambda-\alpha_1)W_N(\lambda-\alpha_2)$$

$$\cdot H_T(\alpha_1-\gamma_1)H_T(-\alpha_1-\gamma_2)$$

$$\left.\cdot\, H_T(\alpha_2-\gamma_3)H_T(-\alpha_2+\gamma_1+\gamma_2+\gamma_3)d\gamma d\alpha\right]d\lambda. \tag{A.17}$$

With (A.1) the first term is equal to

$$\{2\pi H_{2,T}\}^{-2}\int_{\Pi^5} W_N(\lambda-\alpha_1)W_N(\lambda-\alpha_2)\psi_T(\alpha,\gamma)d\gamma d\alpha d\lambda \tag{A.18}$$

$$+\{2\pi H_{2,T}\}^{-1}\sum_{\substack{M\subset\{1,\ldots,6\}\\ M\neq\phi}}\int_{\Pi^5}\left\{\prod_{j\in M} b_j\right\} W_N(\lambda-\alpha_1)W_N(\lambda-\alpha_2)\psi_T(\alpha,\gamma)d\gamma d\alpha d\lambda \tag{A.19}$$

where

$$b_1=\frac{f(\alpha_1)}{f(\lambda)}-1,\quad b_2=\frac{f(\alpha_2)}{f(\lambda)}-1,\quad b_3=\left|\frac{A(\gamma_1)}{A(\alpha_1)}\right|-1,$$

$$b_4=\left|\frac{A(\gamma_1)}{A(\alpha_2)}\right|-1,\quad b_5=\frac{|A(\gamma_2)|}{|A(\alpha_2)|}-1,\quad \text{and}\quad b_6=\left|\frac{A(\gamma_2)}{A(\alpha_1)}\right|-1.$$

Consider a summand of the above sum with $M\cap\{1,2\}\neq\phi$, e.g. with $1\in M$. We have with Theorem 2.4(b)

$$\left(\left|\frac{A(\gamma_1)}{A(\alpha_1)}\right|-1\right)H_T(\alpha_1-\gamma_1)/H_{2,T}^{1/2}$$

$$=O\left[T^{-1/2+\kappa}\sum_{j_i}\sum_{\substack{k_i=1\\ k_1+k_2\geq 1}}^{s_i} T^{-k}L_T(\alpha_1-\gamma_1)^{k+1-k_1-k_2}\right.$$

$$\left.\cdot\, L_T(\alpha_1-\lambda_{1j_1})^{k_1}L_T(\gamma_1-\lambda_{2j_2})^{k_2}\right]$$

and therefore with $L_T(\lambda) \leqq T$

$$\sup(|b_3|, 1)|H_T(\alpha_1-\gamma_1)|/H_{2,T}^{1/2} \leqq KT^{-1/2+\kappa}\left(\sum_{j_i} T^{-1} L_T(\alpha_1-\lambda_{1j_1}) L_T(\gamma_1-\lambda_{2j_2}) + L_T(\alpha_1-\gamma_1)\right).$$

Similarly, we obtain

$$\sup(|b_4|, 1)|H_T(\gamma_1-\alpha_2)|/H_{2,T}^{1/2} \leqq KT^{-1/2+\kappa}\left(\sum_{j_i} T^{-1} L_T(\alpha_2-\lambda_{1j_1}) L_T(\gamma_1-\lambda_{2j_2}) + L_T(\gamma_1-\alpha_2)\right)$$

$$\sup(|b_5|, 1)|H_T(\alpha_2-\gamma_2)|/H_{2,T}^{1/2} \leqq KT^{-1/2+\kappa}\left(\sum_{j_i} T^{-1} L_T(\alpha_2-\lambda_{1j_1}) L_T(\gamma_2-\lambda_{2j_2}) + L_T(\alpha_2-\gamma_2)\right)$$

$$\sup(|b_6|, 1)|H_T(\gamma_2-\alpha_1)|/H_{2,T}^{1/2} \leqq KT^{-1/2+\kappa}\left(\sum_{j_i} T^{-1} L_T(\alpha_1-\lambda_{1j_1}) L_T(\gamma_2-\lambda_{2j_2}) + L_T(\alpha_1-\gamma_2)\right)$$

which implies with Lemma A.1

$$H_{2,T}^{-2} \int_{\Pi^2} \left\{\prod_{j=3}^{6} \sup(|b_j|, 1)\right\} |\psi_T(\alpha, \gamma)| d\gamma \leqq KT^{-2+4\kappa} \log^2 T \sum_{j_i} [L_T(\alpha_1-\alpha_2)^2 + L_T(\alpha_1-\lambda_{1j_1})^2 + L_T(\alpha_2-\lambda_{1j_1})^2]$$

Therefore, the corresponding summand of (A.19) is bounded by

$$KT^{4\kappa-2}\log^2 T \int_{\Pi^3}\Bigg[\sum_{j_i}\sum_{\substack{k_i=0\\k_1+k_2\geqq 1}}^{2s_i} |\alpha_1-\lambda|^{k_1+k_2} \cdot L_{T_{1j_1}}(\lambda-\lambda_{1j_1})^{k_1} L_{T_{2j_2}}(\alpha_1-\lambda_{2j_2})^{k_2} W_N(\alpha_1-\lambda)$$
$$\cdot \sum_{j_i}\sum_{\substack{\ell_i=0\\ \ell_1+\ell_2\geqq 0}}^{2s_i} |\alpha_2-\lambda|^{\ell_1+\ell_2} L_{T_{1j_3}}(\lambda-\lambda_{1j_3})^{\ell_1} L_{T_{2j_4}}(\alpha_2-\lambda_{2j_4})^{\ell_2} W_N(\alpha_2-\lambda)$$
$$\cdot \sum_{j_5}\{L_T(\alpha_1-\alpha_2)^2 + L_T(\alpha_1-\lambda_{1j_5})^2 + L_T(\alpha_2-\lambda_{1j_5})^2\}\Bigg] d\alpha d\lambda$$

Lemma A.3(b) and (c) ($p=10$) now leads to the upper bound

$$KT^{4\kappa-2}\log^2 T\cdot T^{1+1/p}\cdot \int_{\Pi} N^{1-1/p}\Bigg\{\sum_{j_i}\sum_{\substack{k_i=0\\k_1+k_2\geqq 1}}^{2s_i} \frac{L_{T_{1j_1}}(\lambda-\lambda_{1j_1})^{k_1}}{N^{k_1}}\,\frac{L_{T_{2j_2}}(\lambda-\lambda_{2j_2})^{k_2}}{N^{k_2}}$$
$$\cdot \sum_{j_i}\sum_{\ell_i=0}^{2s_i} \frac{L_{T_{1j_3}}(\lambda-\lambda_{1j_3})^{\ell_1}}{N^{\ell_1}}\,\frac{L_{T_{2j_4}}(\lambda-\lambda_{2j_4})^{\ell_2}}{N^{\ell_2}}\Bigg\} d\lambda$$

By using (A.15) this expression is bounded by

$$\leqq KT^{4\kappa-1+1/p}\sum_{i,j}\int_{-\pi}^{\pi} L_T(\lambda-\lambda_{ij})d\lambda=o(T^{-4/5}).$$

Consider now a summand with $M\cap\{1,2\}=\phi$ and $M\cap\{3,4,5,6\}\neq\phi$, e.g. $3\in M$. We obtain in the same way

$$|b_3 H_T(\alpha_1-\gamma_1)|/H_{2,T}^{1/2}\leqq KT^{-1/2+\kappa}\left(\sum_{j_i} T^{-1}L_T(\alpha_1-\lambda_{1j_1})L_T(\gamma_1-\lambda_{2j_2})\right)$$

and therefore

$$H_{2,T}^{-2}\int_{\Pi^2}|b_3|\left\{\prod_{j=4}^{6}\sup(|b_j|,1)\right\}|\psi_T(\alpha,\gamma)|d\gamma$$
$$\leqq KT^{-2+4\kappa}\log^2 T\sum_{j_i}[L_T(\alpha_1-\alpha_2)L_T(\alpha_1-\lambda_{1j_1})+L_T(\alpha_1-\lambda_{1j_1})L_T(\alpha_2-\lambda_{1j_1})$$
$$+L_T(\alpha_1-\lambda_{1j_1})L_T(\alpha_2-\lambda_{2j_2})].$$

Integration over α_1 and α_2 in the corresponding summand in (A.19) gives with $|W_N(\alpha)|\leqq KN$, (3.6) and Lemma A.1 as an upper bound

$$KT^{4\kappa-2}\log^4 T\int_{-\pi}^{\pi} N^2 d\lambda\leqq KT^{4\kappa-1}\log^4 T=o(T^{-4/5}).$$

Therefore, (A.19) is of order $o(T^{-4/5})$. (A.18) is equal to

$$H_{2,T}^{-2}\int_{\Pi^3} W_N(\lambda-\alpha_1)W_N(\lambda-\alpha_2)|H_2^{(T)}(\alpha_1-\alpha_2)|^2 d\alpha d\lambda$$
$$=2\pi H_{4,T}/H_{2,T}^2\int_{\Pi^2} W_N(\lambda-\alpha)^2 d\alpha d\lambda+R \tag{A.20}$$

with

$$|R|\leqq KT^{4\kappa-2}\log^2 T\int_{\Pi^3} W_N(\lambda-\alpha_1)|W_N(\lambda-\alpha_2)-W_N(\lambda-\alpha_1)|L_T(\alpha_1-\alpha_2)^2 d\alpha d\lambda$$
$$\leqq KT^{4\kappa-2}\log^2 T\int_{\Pi^3} N|W_N(\lambda-\alpha_2-\alpha_1)-W_N(\lambda-\alpha_1)|L_T(\alpha_2)^2 d\alpha d\lambda.$$

Since W is of bounded variation we obtain

$$\int_{\Pi}|W_N(\lambda-\alpha_2-\alpha_1)-W_N(\lambda-\alpha_1)|d\alpha_1\leqq KN|\alpha_2|,$$

i.e. the whole expression is bounded by $KT^{4\kappa-2}\log^3 T\int N^2 d\lambda=o(T^{-4/5})$. (A.18) therefore is equal to

$$(2\pi)\left(\lim_{T\to\infty}\frac{TH_{4,T}}{H_{2,T}^2}\right)\int W(\alpha)^2 d\alpha\int_{-\pi}^{\pi}\frac{N}{T}d\lambda+o(T^{-4/5}).$$

By the same methods we obtain that the second summand of (A.17) is of order $o(T^{-4/5})$, and, by using Theorem 2.4(c), that the third summand of (A.17) is also of the same order. (c) is an immediate consequence of (a) amd (b).

The proofs for $i=2$ and for SBIAS, SVAR, and SMSE are completely analogue to the proof for $i=1$. For example in (A.11) to (A.13) we have sums over Fourier-frequencies instead of the β-integrals. We can make use of Lemma A.1(i) to replace these sums by the corresponding integrals in the estimations.

A.5 Properties of AR(s)- and MA(s)-processes

The properties of nontapered estimates and of estimates with a global bandwith are proved in Sect. 4 by considering AR(s)- and MA(s)-processes. We now derive some properties for these processes. In the next section we study the function $H_T(\alpha)$ in the nontapered case.

(A.4) Lemma. *Let X_t be the Gaussian* AR(s)*-process with s-times the characteristic root $p_T=1-1/T$, innovation variance* 1 *and spectral density f. Then*

(a) $X_t \in \mathscr{X}(T, 0, S, \delta_0, C_0)$ *if* $0<\delta_0<\frac{\pi}{3}$ *and* $C_0 \geqq 20$.

(b) $$\frac{f(\lambda+\alpha)}{f(\lambda)}-1=2\pi\{2_{p_T}\}^{2s}\sin^{2s}\left(\frac{\alpha}{2}\right)f(\lambda+\alpha)+O\left(\sum_{j=1}^{2s-1}|\alpha|^j L_T(\lambda+\alpha)^j\right)$$

(c) *If $I_T(\lambda)$ is the periodogram with a data taper of degree (k,κ) where $k=s-1$, then*

$$\int_{-\pi}^{\pi}\left(\frac{EI_T(\lambda)}{f(\lambda)}-1\right)d\lambda=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left(\frac{f(\lambda+\alpha)}{f(\lambda)}-1\right)K_T(\alpha)\,d\alpha\, d\lambda$$

is bounded from below.

Proof. (a) follows from Theorem 2.3. To prove (b) we note that

$$\frac{1-pe^{i\lambda}}{1-pe^{i(\lambda+\alpha)}}=1+pe^{i\lambda}\frac{e^{i\alpha}-1}{1-pe^{i(\lambda+\alpha)}} \tag{A.21}$$

since $|1-p_T e^{i\gamma}|^2 \leqq KL_T(\gamma)^2$ this implies (b). Assertion (c) is derived in the proof of Theorem 7.1 of Dahlhaus (1988). Note, that $E\hat{\theta}_T-\theta_0$ considered in the cited theorem is up to a constant equal to $\int_{-\pi}^{\pi}\left(\frac{EI_T(\lambda)}{f(\lambda)}-1\right)d\lambda$.

A.6 Elementary Properties in the Nontapered Case

Let

$$\Delta_T(\alpha):=\sum_{t=0}^{T-1}\exp(-i\alpha t)=\frac{\exp(-i\alpha T)-1}{\exp(-i\alpha)-1}=\exp\left(-i\alpha\frac{T-1}{2}\right)\frac{\sin(T\alpha/2)}{\sin(\alpha/2)}$$
$$=H_T(\alpha) \quad \text{if} \quad h_{t,T}=1 \;(t=0,\dots,T-1) \tag{A.22}$$

(A.5) Lemma. *Let $h_{t,T}=1$ $(t=0,\dots,T-1)$. Then*
(a) $H_{k,T}=T$ *for all* $k\in\mathbb{N}$,
(b) $|\Delta_T(\alpha)|\leqq KL_T(\alpha)$ *with a* $K>0$,
(c) $|\Delta_T(\alpha)|\geqq KT$ *for all* $|\alpha|\leqq\pi/T$ *with a* $K>0$
(d) *If* $p_T=1-1/T$, *then*

$$\frac{1-p_T e^{i\alpha}}{1-p_T e^{i\gamma}}\,\Delta_T(\gamma-\alpha)=\Delta_T(\gamma-\alpha)+p_T e^{i\gamma}\,\frac{1-e^{-i(\gamma-\alpha)T}}{1-pe^{i\gamma}}$$
$$=O\,[L_T(\gamma-\alpha)+L_T(\gamma)]\,. \tag{A.23}$$

Proof. All proofs are straightforward. (d) follows with (A.21) and (A.22).

A.7 Global Bandwith Selection

For the proof of Theorem 4.1 we need the following lemma. It is the analogue result to Lemma A.3, now for a global bandwith.

(A.6) Lemma. *Let W_N be defined as in Lemma A.3 and $N \leqq T$. Then we have for all $k \in \mathbb{N}_0$ and $\ell \in \mathbb{N}$*

$$\int_{-\pi}^{\pi} |\alpha|^k L_T(\lambda+\alpha)^\ell W_N(\alpha)\,d\alpha \leqq K\,\frac{L_N(\lambda)^\ell}{N^k}\left\{\left(\frac{T}{N}\right)^{\ell-1}+\log T\right\}.$$

Proof. By considering the cases $|\lambda| \leqq N^{-1}$ and $|\lambda| > N^{-1}$ we obtain the result with Lemma A.3(a) and Lemma A.1(b), (f).

Proof of Theorem 4.1. (a.1) Let $i=1$ and $s_1=0$. Similarly to (A.11)–(A.13) we have

$$\frac{Ef_T^{(1)}(\lambda)}{f(\lambda)}-1=\int_{-\pi}^{\pi}\left[\frac{f(\lambda+\alpha)}{f(\lambda)}-1\right]K_T(\alpha)\int_{-\pi}^{\pi}\left[\frac{f(\lambda+\alpha+\beta)}{f(\lambda+\alpha)}-1\right]W_N(\beta)\,d\beta\,d\alpha$$
$$+\int_{-\pi}^{\pi}K_T(\alpha)\int_{-\pi}^{\pi}\left[\frac{f(\lambda+\alpha+\beta)}{f(\lambda+\alpha)}-1\right]W_N(\beta)\,d\beta\,d\alpha$$
$$+\int_{-\pi}^{\pi}\left[\frac{f(\lambda+\alpha)}{f(\lambda)}-1\right]K_T(\alpha)\,d\alpha\,. \tag{A.24}$$

Analogously to (A.14) we get

$$\left|\frac{f(\lambda+\alpha)}{f(\lambda)}-1\right|K_T(\alpha)\leqq KT^{2\kappa-1}\sum_{j_2}L_T(\lambda+\alpha-\lambda_{2j_2})L_T(\alpha)\,.$$

Thus, the λ-integral over the third term is $O\,[T^{2\kappa-1}\log^2 T]=o\left[\frac{T^{2s_2-1}}{N^{2s_2}}\right]$. We obtain with Theorem 2.4(a) ($\ell=1, m_1=0$) and Lemma A.6

$$\int_{-\pi}^{\pi}\left[\frac{f(\lambda+\alpha+\beta)}{f(\lambda+\beta)}-1\right]W_N(\beta)\,d\beta\leqq K\left(\frac{T}{N}\right)^{2s_2-1}.$$

Therefore, the λ-integral over the first term of (A.24) is

$$O\left[T^{2\kappa}\log^2 T\,\frac{T^{2s_2-2}}{N^{2s_2-1}}\right]=o\left(\frac{T^{2s_2-1}}{N^{2s_2}}\right).$$

The λ-integral over the second term of (A.24) is equal to

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\frac{f(\lambda+\beta)}{f(\lambda)}-1\right]W_N(\beta)\,d\beta\,d\lambda\,. \tag{A.25}$$

We now consider the spectral density of the Gaussian AR(s_2)-process with s_2 times the characteristic root $p_T = 1 - 1/T$ and innovation variance 1. By using Lemma A.4(b) and Lemma A.6 (A.25) is equal to

$$2\pi\{2p_T\}^{s_2} \int_{-\pi}^{\pi} f(\lambda)d\lambda \int_{-\pi}^{\pi} \sin^{s_2}\left(\frac{\beta}{2}\right) W_N(\beta)d\beta + O\left(\frac{T^{2s_2-2}}{N^{2s_2-1}}\right).$$

Since $\int_{-\pi}^{\pi} f(\lambda)d\lambda \geqq cT^{2s_2-1}$ this is bounded form below by $c\,\frac{T^{2s_2-1}}{N^{2s_2}}$ with some constant $c > 0$. The case $i = 2$ is proved analogously. If $s_2 = 0$ we consider in the last part of the proof a MA(s_1)-process with s_1 times the characteristic root $p_T = 1 - 1/T$ and obtain the same result as before. (a.2) is an immediate consequence of (a.1) (Cauchy-Schwarz inequality). Furthermore, we know for a process with fixed spectral density (independent of T) that $\mathrm{IVAR}(f_T^{(i)}) \geqq c\,\frac{N}{T}$. This implies $\sup_{\mathscr{X}_T} \mathrm{IMSE}(f_T^{(i)}) \geqq c \max\left(\frac{T^{4s_2-2}}{N^{4s_2}}, \frac{N}{T}\right) \geqq cT^{-2/(4s_2+1)}$.

(b) can be checked by a straightforward modification of the proof of Theorem 3.6 by using Lemma A.6 instead of Lemma A.3(b). We omit details.

A.8 The Leakage and the Trough Effect: Proof of Theorem 4.3

Proof of Theorem 4.3. Let $i = 1$ and $(s_1 \leqq 1,\ s_2 > 0)$. Then $k < s_2$. Let X_t be the Gaussian AR(s)-process with $s = k + 1$ times the characteristic root $p_T = 1 - 1/T$. Theorem 2.3 implies $X_t \in \mathscr{X}(T, s_1, s_2, \delta_0, C_0)$. We now prove that the λ-integrals over (A.11) and (A.13) tend to zero while the λ-integral over (A.12) is bounded from below. We get with Theorem 2.4(a) ($\ell = 1$), Lemma 3.3 and Lemma A.1

$$\left|\int_{-\pi}^{\pi} \left(\frac{f(\lambda+\alpha+\beta)}{f(\lambda+\beta)} - 1\right) K_T(\alpha)d\alpha\right|$$

$$\leqq KT^{-2s+2\kappa+1} \sum_{j=1}^{2s} \int_{-\pi}^{\pi} L_T(\lambda+\alpha+\beta)^j L_T(\alpha)^{2s-j} d\alpha$$

$$\leqq KT^{2\kappa} \log T.$$

Furthermore, we obtain with Lemma A.3 or Lemma A.6

$$\int \int_{-\pi}^{\pi} \left|\frac{f(\lambda+\beta)}{f(\lambda)} - 1\right| W_N(\beta)d\beta d\lambda \leqq KT^{-1/5}$$

which implies that the λ-integrals over (A.11) and (A.13) tend to zero. The λ-integral over (A.12) is equal to

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left[\frac{f(\lambda+\alpha)}{f(\lambda)} - 1\right] K_T(\alpha)d\alpha d\lambda$$

which is bounded from below by Lemma A.4(c). This prove (a) for $(s_1 \leqq 1, s_2 > 0)$.

If $(s_1 > 0, s_2 \leqq 1)$ we use instead the representation (A.24) and consider an MA(s)-process with $s = k + 1$ times the characteristic root $p_T = 1 - 1/T$. Similarly, we obtain

$$\left| \int_{-\pi}^{\pi} \left(\frac{f(\lambda+\alpha+\beta)}{f(\lambda+\alpha)} - 1 \right) W_N(\beta) d\beta \right| \leqq K \sum_{j=1}^{2s} N^{-j} L_T(\lambda+\alpha)^j$$

and therefore for the λ-integral over the first term in (A.24) as an upper bound

$$KT^{2\kappa-2s+1} \sum_{j,k=1}^{2s} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} N^{-j} L_T(\lambda)^k L_T(\alpha)^{2s-k} L_T(\lambda+\alpha)^j \, d\alpha \, d\lambda$$

We obtain as an upper bound with Lemma A.1(f), (g)

$$KT^{2\kappa-2s+1} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{j=1}^{2s} \sum_{k=1}^{2s-1} N^{-j} [L_T(\lambda)^{2s} L_T(\lambda+\alpha)^j + L_T(\lambda)^{k+j} L_T(\alpha)^{2s-k}] \, d\alpha \, d\lambda$$

If N is local i.e. $N \geqq cT^{1/5} L_T(\lambda)^{4/5}$ from (3.7), this is with Lemma A.1(b) bounded by $KT^{2\kappa-4/5}$. If N is global this is bounded by $KT^{2\kappa+2s-1}/N^{2s}$. Thus, the λ-integral over the first term of (A.24) tends to zero.

The λ-integral over the second term is equal to

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left(\frac{f(\lambda+\beta)}{f(\lambda)} - 1 \right) W_N(\beta) \, d\beta \, d\lambda = o(1).$$

The λ-integral over the third term is equal to

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left(\frac{f(\lambda)}{f(\lambda+\alpha)} - 1 \right) K_T(\alpha) \, d\alpha \, d\lambda = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left(\frac{f^{-1}(\lambda+\alpha)}{f^{-1}(\lambda)} - 1 \right) K_T(\alpha) \, d\alpha \, d\lambda$$

which is bounded from below by Lemma A.4(c) (since $f^{-1}(\lambda)$ is the spectral density of an AR(s)-process with root $\rho_T = 1 - 1/T$). This proves (a) for ($s_1 > 0, s_2 \leqq 1$).

(b) and (c) are immediate consequences. The case $i = 2$ and the results for SBIAS and SMSE are proved analogously.

Proof of Corollary 4.4. Suppose e.g. $s_1 \geqq 0$. Since the nontapered case $h_{t,T} \equiv 1$ is a data taper of degree (0,0) we obtain the assertion from the relation

$$\mathscr{X}(T, s_1, s_2, \delta_0, C_0) \supset \mathscr{X}(T, s_1, 0, \delta_0, C_0).$$

A.9 The Variance Effect: Proof of Theorem 4.5

Proof of Theorem 4.5. Let $s_2 > 0$ and $X_t \in \mathscr{X}_T$ be the Gaussian AR(1)-process with the characteristic root $\rho_T = 1 - 1/T$ and innovation variance 1. Again we have relation (A.17) with $f_4(\gamma) = 0$. The Cauchy-Schwarz inequality, Theorem 2.4(a) ($\ell = 0$), Lemma A.5 and Lemma A.1(e) imply

$$\begin{aligned} & H_{2,T}^{-2} \int_{\Pi^2} \frac{f(\gamma_1) f(\gamma_2)}{f(\alpha_1) f(\alpha_2)} \psi_T(\alpha, \gamma) \, d\gamma \\ & = T^{-2} \prod_{i,j=1}^{2} \left\{ \int_{\Pi} \frac{f(\gamma_i)}{f(\alpha_j)} |H_T(\gamma_i - \alpha_j)|^2 \, d\gamma_i \right\}^{1/2} \\ & \leqq KT^{-2} \prod_{i,j=1}^{2} \left\{ \int_{\Pi} \sum_{k=0}^{2} L_T(\gamma_i)^k L_T(\gamma_i - \alpha_j)^{2-k} \, d\gamma_i \right\}^{1/2} \\ & \leqq K. \end{aligned}$$

Furthermore, we obtain with Theorem 2.4(a) ($\ell=1$), and Lemma A.3 or Lemma A.6

$$\int_{\Pi^3} \left(\frac{f(\alpha_1)f(\alpha_2)}{f(\lambda)^2}-1\right) W_N(\lambda-\alpha_1)\, W_N(\lambda\pm\alpha_2)\,d\alpha_1\, d\alpha_2\, d\lambda = o(1).$$

Thus, we get with (A.17)

$$\int_{-\pi}^{\pi} \operatorname{var}\left(\frac{\hat{f}_T^{(1)}(\lambda)}{f(\lambda)}\right) d\lambda = \{2\pi T\}^{-2} \int_{\Pi^5} W_N(\lambda-\alpha_1)\,[W_N(\lambda-\alpha_2) + W_N(\lambda+\alpha_2)]\,\frac{f(\gamma_1)f(\gamma_2)}{f(\alpha_1)f(\alpha_2)}\,\psi_T(\alpha,\gamma)\,d\gamma\, d\alpha\, d\lambda + o(1)$$

which, by applying (A.1) and Lemma A.5(d), is equal to

$$\begin{aligned} &T^{-2} \int_{\Pi^5} W_N(\lambda-\alpha_1)\,[W_N(\lambda-\alpha_2)+W_N(\lambda+\alpha_2)]\,p_T^4 f(\gamma_1)f(\gamma_2) \\ &\cdot [1-e^{-i(\alpha_1-\gamma_1)T}][1-e^{-i(\gamma_1-\alpha_2)T}][1-e^{-i(\alpha_2-\gamma_2)T}][1-e^{-i(\gamma_2-\alpha_1)T}]\,d\gamma\, d\alpha\, d\lambda \\ &+O\left[T^{-2}\int_{\Pi^5} W_N(\lambda-\alpha_1)\,W_N(\lambda-\alpha_2)\,L_T(\alpha_1-\gamma_1)\,L_T(\gamma_1)\,L_T(\gamma_2)^2\, d\gamma\, d\alpha\, d\lambda\right] \\ &+\text{similar terms } + o(1). \end{aligned} \qquad \text{(A.26)}$$

The O-term is bounded by

$$KT^{-1}\log T \int_{\Pi^3} N W_N(\lambda-\alpha_2)\,L_T(\alpha_1)\,d\alpha\, d\lambda = o(1).$$

The same holds for all other 'similar terms'. We have

$$\int_{-\pi}^{\pi} f(\gamma)\exp(i\alpha u)\,d\alpha = \frac{p_T^{|u|}}{1-p_T^2}.$$

Let $w_N(u) = \int_{-\pi}^{\pi} W_N(\alpha)\exp(i\alpha u)\,d\alpha$. Extensive but straighforward calculations show that (A.26) is equal to

$$\int_{-\pi}^{\pi} 8\pi^2 p_T^4 T^{-2}(1-p_T^2)^{-2}\{(w_N(0)-w_N(T)p_T^T\cos\lambda T)^2 + (p_T^T w_N(0) - w_N(T)\cos\lambda T)^2\}\,d\lambda$$

Since $(1-p_T^2) = \frac{2}{T} - \frac{1}{T^2}$, $w_N(0)=1$, $|w_N(T)| \leqq w_N(0)$, and $p_T^T = (1-1/T)^T \to e^{-1}$ this expression is bounded from below, and Theorem 4.5 is proved. If $s_2=0$ and $s_1>0$ we consider instead the Gaussian MA(1)-process with characteristic root $p_T = 1-1/T$. The result then follows analogously.

A.10 Segment Estimates: Proof of Theorem 5.1

Proof. Following the proof of Theorem 2.2 in Dahlhaus (1985) we get for a Gaussian process

$$\int_{-\pi}^{\pi} \operatorname{var}\left(\frac{f_T^{(3)}(\lambda)}{f(\lambda)}\right) d\lambda = \{2\pi M H_{2,N}\}^{-2} \left[\int_{\Pi^3} \frac{f(\gamma_1) f(\gamma_2)}{f(\lambda) f(\lambda)} |H_N(\lambda-\gamma_1)|^2 \right.$$
$$\cdot |H_N(\lambda-\gamma_2)|^2 |\Delta_M(L\gamma_1 - L\gamma_2)|^2 d\gamma_1 d\gamma_2 d\lambda$$
$$\left. + \sum_{s,t=1}^{M} \int_{\Pi} \left| \int_{\Pi} \frac{f(\gamma_1)}{f(\lambda)} H_N(\lambda-\gamma_1) H_N(\lambda+\gamma_1) e^{-iL\gamma_1(s-t)} d\gamma_1 \right|^2 d\lambda \right]. \tag{A.27}$$

Both summands are positive. We prove the lower bound for the first summand. Let $X_T \in \mathscr{X}_T$ be the Gaussian AR(1)-process with characteristic root $p_T = 1 - 1/T$ and innovation variance 1. Let

$$H_N^D(\alpha) = \sum_{t=-1}^{N-1} \{h_{t+1,N} - h_{t,N}\} \exp(-i\alpha t) \quad \text{with} \quad h_{N,N} = h_{-1,N} = 0.$$

Summation by parts yields

$$H_N(\alpha) = \{\exp(i\alpha) - 1\}^{-1} H_N^D(\alpha). \tag{A.28}$$

With Lemma 3.3, (A.1) and Lemma A.4(b) therefore the first summand of (A.27) is equal to

$$\{2\pi M H_{2,N}\}^{-2} \left[\int_{\Pi^3} |H_N(\lambda-\gamma_1)|^2 |H_N(\lambda-\gamma_2)|^2 |\Delta_M(L\gamma_1 - L\gamma_2)|^2 d\gamma d\lambda \right.$$
$$\left. + 4\pi^2 p_T^2 \int_{\Pi^3} f(\gamma_1) f(\gamma_2) |H_N^D(\lambda-\gamma_1)|^2 |H_N^D(\lambda-\gamma_2)|^2 |\Delta_M(L\gamma_1 - L\gamma_2)|^2 d\gamma d\lambda \right]$$
$$+ O\left[N^{4\kappa-4k-2} \sum_{\substack{k_1,k_2=0 \\ 1 \leq k_1+k_2 \leq 3}}^{2} \int_{\Pi^3} L_T(\gamma_1)^{k_1} L_N(\lambda-\gamma_1)^{2k+2-k_1} L_T(\gamma_2)^{k_2} \right.$$
$$\left. \cdot L_N(\lambda-\gamma_2)^{2k+2-k_2} d\gamma \right] \tag{A.29}$$

By using Lemma A.1 we get as an upper bound for the 0-term

$$K N^{4\kappa-2} T \log T \log N$$

which is $o\left(\frac{T^2}{N^3}\right)$ in the tapered and $o\left(\frac{T^2}{N^2}\right)$ in the nontapered case. By using the definition of H_N and Δ_M we obtain that the first term of (A.29) is larger than $C\frac{N}{ML} \geqq C\frac{N}{T}$ if $L \leqq N$ and larger than $C\frac{1}{M} \geqq C\frac{N}{T}$ if $L > N$. Since $|\gamma_i| \leqq (2T)^{-1}$ implies $|L\gamma_1 - L\gamma_2| \leqq \frac{\pi}{M}$ (both, for $L \leqq N$ and $L > N$) we obtain with Lemma A.5(c) and (2.6) as an lower bound for the second term of (A.29)

$$CT^4 H_{2,N}^{-2} \int_{\Pi} \left\{ \int_{|\gamma| \leqq (2T)^{-1}} |H_N^D(\lambda-\gamma)|^2 d\gamma \right\}^2 d\lambda . \tag{A.30}$$

Suppose now that no data taper is applied. Then

$$H_N^D(\alpha) = \exp(i\alpha) - \exp(-i\alpha(N-1)) .$$

Therefore, we obtain with $\kappa = 0$

$$\{|H_N^D(\lambda-\gamma)|^2 - |H_N^D(\lambda)|^2\}/H_{2,N} \leqq KN^{2\kappa}|\gamma| \tag{A.31}$$

and (A.30) is larger than

$$CT^2 H_{2,N}^{-2} \int_{\Pi} |H_N^D(\lambda)|^4 d\lambda + O\left[T^4 H_{2,N}^{-1} \int_{\Pi} \int_{|\gamma_1| \leqq (2T)^{-1}} \{|H_N^D(\lambda-\gamma_1)|^2 + |H_N^D(\lambda)|^2\} \int_{|\gamma_2| \leqq (2T)^{-1}} N^{2\kappa} |\gamma_2| d\gamma d\lambda \right]$$

$$\geqq C \frac{T^2}{N^2} + O\left(\frac{T}{N}\right) \geqq C \frac{T^2}{N^2} \geqq C \tag{A.32}$$

If a taper is applied we again obtain (A.31) and (A.32). (A.28), Lemma 3.3, and $H_{2,N} \sim N$ imply

$$|H_N^D(\alpha)|/H_{2,N}^{1/2} \leqq KN^{-k-1/2+\kappa} L_N(\alpha)^k \leqq K^{-3/2+\kappa} L_N(\alpha) . \tag{A.33}$$

Therefore, by using Lemma A.1 the O-term is bounded by

$$KTN^{-2+4\kappa} = o(T^2 N^{-3-4\kappa}) .$$

The Hölder inequality implies

$$2\pi \sum_{t=-1}^{N-1} \{h_{t+1,N} - h_{t,N}\}^2 = \int_{\Pi} |H_N^D(\lambda)|^2 d\lambda$$

$$\leqq KN^{-1+\kappa} \left\{ \int_{\Pi} |H_N^D(\lambda)|^4 d\lambda \right\}^{1/4} \left\{ \int_{\Pi} L_N(\lambda)^{4/3} d\lambda \right\}^{3/4}$$

$$\leqq KN^{-3/4+\kappa} .$$

Straighforward calculations give (cp. Def. 2.3)

$$2\pi \sum_{t=-1}^{N-1} \{h_{t+1,N} - h_{t,N}\}^2 = \frac{2\pi}{N} \int_0^1 h'(x)^2 dx + O\left[\frac{N^{2\kappa}}{N^2}\right] \geqq CN^{-1}$$

and therefore as a lower bound for (A.32) and for the second term of (A.29)

$$CT^2 N^{-3-4\kappa} \quad \text{with some constant} \quad C > 0 .$$

Together with the lower bound $C \frac{N}{T}$ for the first term of (A.29) this leads to a convergence rate of at most $T^{-(1+4\kappa)/(4+4\kappa)}$.

The proof for the Kolmogorov-Zhurbenko taper is the same as for a taper of degree (k, κ). In particular we also have (A.31) and (A.33). We omit details.

If $s_2 = 0$ we consider instead a Gaussian MA(1)-process with characteristic root $p_T = 1 - 1/T$. This leads to the same result. We omit the proof of this case.

References

Bartlett, M.S.: Periodogram analysis and continuous spectra. Biometrika **37**, 1–16 (1950)

Bloomfield, P.: Fourier analysis of time series: an introduction. New York: Wiley Interscience 1976

Brillinger, D.R.: Time series: data analysis and theory. San Francisco: Holden-Day 1981

Childers, G. (ed.): Modern Spectrum Analysis. New York: IEEE 1978

Dahlhaus, R.: Spectral analysis with tapered data. J. Time Ser. Anal. **4**, 163–175 (1983)

Dahlhaus, R.: Small sample effects in time series analysis: a new asymptotic theory and a new estimate. Ann. Stat. **16**, 808–841 (1988)

Franke, J., Härdle, W.: On bootstrapping kernel spectral estimates. Universität Kaiserslautern (preprint) 1988

Hannan, E.J.: Multiple time series. New York: Wiley 1970

Kolmogorov, A.N., Zhurbenko, I.G.: Estimations of spectral functions of stochastic processes. Proc. 11th European Meeting of Statisticians, Oslo, 36. 1978

Priestley, M.B.: Spectral analysis and time series, vol. 1. London: Academic Press 1981

Tukey, J.W.: An introduction to the calculations of numerical spectrum analysis. In: Harris, B. (ed.) Advanced seminar on spectral analysis of time series. pp 25–46. New York: Wiley 1967

Welch, P.D.: The use of the Fast Fourier Transform for estimation of spectra: a method based on time averaging over short, modified periodograms. IEEE Trans. Audio Electroacoustics AU-**15**, 70–74 (1967)

Zhurbenko, I.G.: On the efficiency of spectral density estimates of a stationary process. Theory Probab. Appl. **25**, 466–480 (1980)

Zhurbenko, I.G.: On the efficiency of spectral density estimates of a stationary process, II. Theory Probab. Appl. **28**, 409–419 (1983)